LIBRARY OF PRINCETON
AUG 18 1944
THEOLOGICAL SEMINARY

HN29
.L96

SOCIAL RESEARCH

SOCIAL RESEARCH

A Study in Methods of Gathering Data

LIBRARY OF PRINCETON
AUG 18 1944
THEOLOGICAL SEMINARY

BY

GEORGE A. LUNDBERG

BENNINGTON COLLEGE

LONGMANS, GREEN AND CO.

NEW YORK · LONDON · TORONTO

1942

LONGMANS, GREEN AND CO.
55 FIFTH AVENUE, NEW YORK
221 EAST 20TH STREET, CHICAGO

LONGMANS, GREEN AND CO., LTD.
OF PATERNOSTER ROW
43, ALBERT DRIVE, LONDON, S.W.19
17, CHITTARANJAN AVENUE, CALCUTTA
NICOL ROAD, BOMBAY
36A, MOUNT ROAD, MADRAS

LONGMANS, GREEN AND CO.
215 VICTORIA STREET, TORONTO

LUNDBERG
SOCIAL RESEARCH

COPYRIGHT · 1942
BY LONGMANS, GREEN AND CO., INC.
ALL RIGHTS RESERVED, INCLUDING THE
RIGHT TO REPRODUCE THIS BOOK, OR
ANY PORTION THEREOF, IN ANY FORM

SECOND EDITION

PRINTED IN THE UNITED STATES OF AMERICA

To

JOHN M. GILLETTE

AND TO

STUDENTS OF THE FIRST EDITION

"The truth is, the Science of Nature has been already too long made only a work of the *Brain* and the *Fancy :* It is now high time that it should return to the plainness and soundness of *Observations* on *material* and obvious things." Robert Hooke, *Micrographia,* 1665, Preface.

"The present tendency is to loaf and to generalize. . . We need men not afraid to work : who will get busy the adding machine and the logarithms and give us *exact studies,* such as we get from the psychological laboratories. *Sociology can be made an exact, quantitative science* if we can get *industrious* men interested in it." F. H. Giddings, *Amer. Jour. Sociol.,* 15:196, 1909–10.

"The critical habit of thought, if usual in a society, will pervade all its mores, because it is a way of taking up the problems of life. Men educated in it cannot be stampeded by stump orators and are never deceived by dithyrambic oratory. They are slow to believe. . ." W. G. Sumner, *Folkways,* 1907, pp. 632-33.

"The *fundamentals* of the science which our century inherited from its predecessors have been *modified,* now slightly, now quite perceptibly, till our outlook on the physical universe today bears but little resemblance to that of only thirty years ago. Great and striking as these advances are, there has been another, most rapidly developed since 1930, which has been slowly gathering momentum for all of 2300 years, which is of far deeper significance for "truth" — or Pilate's query — than any of the radical advances of science of the past thirty years. Being more fundamental, more radical, and simpler than any of the spectacular advances in science, naturally this new advance has escaped the notice which its far-reaching importance merits. Yet it is of profound significance for all theorizing and truth-seeking, scientific or other." — E. T. Bell, *The Search for Truth* (Reynal and Hitchcock, 1934), pp. 7-8.

"Some of the gratuitous dualisms done away with [by modern biology, cultural anthropology, and the experimental method] . . . are those of the objective and subjective, the real and apparent, the mental and physical, scientific physical objects and objects of perception, things of experience and things-in-themselves concealed behind experience, the latter being an impenetrable veil which prevents cognitive access to the things of nature." — John Dewey, "Experience, Knowledge, and Value : A Rejoinder." *(The Philosophy of John Dewey,* P. A. Schilpp, Editor, Northwestern University Press, 1939), p. 524.

PREFACE

TO THE FIRST EDITION

The advance of physical science has been chiefly dependent upon the refinement of the technique and tools of observation and description. In the social sciences this technique and these tools are as yet largely undeveloped. There is a large number of treatises on "Principles" of Sociology, Economics, and Political Science. As hypotheses, these Principles are of unquestionable value. But the hope for the advancement of the social sciences lies in the testing of these principles by a patient accumulation of observed data, scientifically classified and generalized. The technique of such observation and classification is today the aspect of the social studies which is most in need of development.

To be sure, the details of an adequate technology and methodology in a given case can be developed only as we work on specified problems. "The thought of an end," says John Dewey, "is strictly correlative to perception of means and methods. Only when and as the latter become clear during the serial process of execution, does the project and guiding aim and plan become evident and articulated." It would certainly be absurd to attempt to prescribe a rigid method of procedure to be followed in all inquiries. However, there are certain principles of logic, certain common dangers of bias, and certain common general approaches, which may be treated as generally present in all research. It is with these general principles and methods in the social sciences that this book is concerned.

The present work has been produced primarily in order to bring together and to make available in compact form materials which the author has found useful in his course on Methods of Social Research. The objectives of this course, and hence of the present book, are:

1. To emphasize the importance of accurate and objective observation as the first step in scientific method, on which the value of subsequent analysis and the validity of the conclusions depend.

2. To inculcate a healthy skepticism of, and a critical spirit toward, statistical data by acquainting the student with the difficulties, danger, and inaccuracies to which the collection and interpretation of social data are especially subject.

3. To give a general knowledge of the technique of gathering original data.

The value of the first two of these objectives to the general student will probably be admitted by everyone. With the increasing use of quantitative data in the literature and social administration of the day, a critical discrimination in their production and consumption becomes increasingly desirable. The last objective, in addition to being a goal in itself to the research student and the field worker, is a means to the achievement of the first two purposes.

The need for a discussion of the topics treated in this book was impressed on the author by the fact that while there are several dozens of treatises on statistics and methods of classifying, generalizing, and scientifically manipulating quantitative social data, there are only three or four which place the main emphasis on the *scientific observation and recording of these data*. The main reason for this fact is, of course, that the methods of handling and manipulating data are largely the same in all sciences, and hence it is natural that this phase of scientific method should be much more highly developed in the social sciences than the first step, namely, the observation and recording of data. This is the step in scientific method which is likely to be most different in the various fields of science, because it is the step concerned largely with the matter of units, nomenclature, and instruments of observation which may be very different in different fields. As a result, statistical technique of manipulating data already collected in the social sciences has become more highly developed than the technique of observing and recording the data.

Obviously there is little to be gained by the application of refined statistical technique to data which are grossly inaccurate, indefinite, or otherwise defective from the standpoint of observation and recording. Says Wilcox: "I cannot but feel that there may be some danger of turning the attention of the class too exclusively either to graphic methods of rep-

resenting results or to mathematical methods of determining the exact degree of correlation. Doubtless each of these has an important place in a course in statistics, but neither strikes at the spot in which most errors are found. . . For a large majority of statistical errors are those which creep into the raw material and cannot be eliminated by any refinements of mathematical analysis or any skill in graphic representation. To center attention on these field errors has the incidental advantage of bringing home to the student as nothing else can the real difficulties before the official statistician. . ."[1] Certainly it is this phase of social science which is in greatest need of development today. To call attention to this need, as well as to make a few general suggestions toward meeting it, is the purpose of the present work.

I have tried to give full credit in the text and in footnotes to the various authors on whose works I have drawn, but realize that it is perhaps impossible to make adequate acknowledgments in this way. Especially is this true of the ideas I have drawn from my former teachers. First among these should be mentioned Professor L. L. Bernard, to whom more than to any other person I am indebted for my general viewpoint as well as for numerous quotations in the text. To Professor F. Stuart Chapin I owe a threefold debt — for training in quantitative methods, for permission to draw liberally on his published works, and for reading a large portion of the manuscript and making valuable suggestions. To Professor M. C. Elmer I am likewise indebted for his reading of a large portion of the manuscript and for constant encouragement and suggestions. I am also under obligation to Professor R. E. Chaddock for valuable suggestions, encouragement, and criticism of part of the manuscript.

In addition to the above I wish to express my appreciation to Professors L. L. Thurstone and Jerome Davis, and to Dr. Emil Frankel for permission to quote liberally from their published writings, and to Dorothy Swaine Thomas for the privilege of examining a portion of the manuscript of "The Child in America" by D. S. and W. I. Thomas. Acknowl-

[1] W. F. Wilcox, "Cooperation between Academic and Official Statisticians," *Publications of the American Statistical Association*, Vol. 14 (N. S.), 1914–15, pp. 292-93.

edgments are also due to the *American Journal of Sociology,* The Century Company, and *Social Forces,* for permission to reprint extracts from their publications.

Finally, I am under great obligation to Professor John F. Markey for reading the major part of the manuscript and making suggestions of great importance ; to my colleague Calvin F. Schmid for a discriminating reading of the proofs ; and above all to Professor Read Bain, who has not only read the entire manuscript critically but has given unsparingly of his time in extensive criticism, much discussion, and constant encouragement. For the errors and shortcomings which remain, the author is, of course, solely responsible.

GEORGE A. LUNDBERG

Pittsburgh, Pa.
July 1, 1929

PREFACE

TO THE SECOND EDITION

The extensive developments in methods of social research during the past decade have necessitated a complete rewriting of all but four chapters of the First Edition. Extensive additions have been made even to those chapters of the First Edition which have been in part retained (Chapters II-V). Occasional paragraphs and pages of other chapters of the original edition have also been incorporated into the present text, but nearly all of these passages have been modified to some degree. The orginal chapter headings have been retained for the most part, with some changes in their order. The chapter on Case Work and the Statistical Method of the First Edition has been eliminated as a separate chapter. The aspects of this subject which are of contemporary interest are discussed in Chapters I and XI of the present text. Two new chapters have been added, namely, Chapter VII on Questionnaires and Chapter X on Sociometric Methods in Ecology and Inter-personal Relations. At least three-fourths of the present edition is entirely new, both as regards the text and the research referred to. Most of the references are to studies that have appeared during the last five years.

The present text places greater emphasis than did the First Edition on practical problems of research, and devotes relatively less space to controversial theoretical topics. Two conditions have influenced me in this modification. First, I have devoted considerable space in another recent volume to the more theoretical questions of methodology. Instead of going over this ground again in the present volume, the position in question has been briefly summarized, and references made to the more complete discussion to be found in my *Foundations of Sociology* (Macmillan, 1939).

Second, my experience in practical non-academic research since the first edition appeared, has doubtless influenced both the selection of subject matter and its treatment in the present edition. For example, my research opportunities as Director of the Bureau of Social Research of the Pitts-

burgh Federation of Social Agencies, as director of a research project on leisure and recreation for the Council for Research in the Social Sciences of Columbia University, and as Research Supervisor in the Federal Emergency Relief Administration, impressed me with certain aspects of research which were relatively neglected in the First Edition. My recent community studies on interpersonal relations, likewise, have caused me to deal in the present edition with some of the problems and techniques of this type of research. In any case, the present text is primarily a textbook on practical techniques of securing basic data on sociological questions. It aims to make available to students and teachers the most valuable recent experience and criticism on this subject. Both for these reasons, and because of the general trend of research during the past decade, the reader will probably find much less material of a controversial nature in the present edition, and very little with which anyone will disagree with reference to practical advice to researchers, regardless of his individual views on a number of questions of sociological theory.

At the same time, the breadth of the field and the variety of the methods prevent giving equally full treatment to all of them. This is not necessarily a reflection on the value of the neglected methods. In view of the fact that this point is frequently overlooked, it may not be amiss to repeat what I said in the First Edition on this subject:

> The social sciences cannot afford, in their present stage of development, at least, to neglect any method which promises to throw light on the problems which we seek to solve. The argument in this book is for the refinement and objectification of all the methods which at present admittedly contribute something to our understanding of social behavior. There is, for example, no desire to discredit or abandon the fullest possible use of life history documents, intensive historical studies of individual cases or institutions, and other discursive and non-quantitative approaches. What has been advocated is that the technique of classifying and analyzing such information be improved in such a way as to permit scientific generalizations from the data. This involves the development of a technique for the statistical summarization of these case studies. These methods tend to merge with the procedures we have considered, to the extent that they become objective in their technique. Instead of regarding life

history and case study methods as fundamentally different from the tests and statistical techniques, therefore, they should be regarded as different stages in the development of a common scientific method. The final test of the validity of a method is its efficacy in securing the type of results sought. It follows that we are justified in the use of any method, regardless of its degree of objectivity, if it helps to throw even a little light on the problem we are seeking to solve. From this point of view most social problems call for a composite method — the full utilization of historical and genetic sources, the non-quantitative description of individual cases in complete social situations, as well as the more rigid, objective, and quantitative tests — all of which may give us a more accurate account than any single approach can give. The purpose in pointing out the weaknesses of certain methods, therefore, is not to advocate their abandonment as long as they contribute something to our knowledge. The purpose of evaluating them has been to help us avoid extravagant assumptions as to the validity of the conclusions which they are capable of yielding. (First Edition, pp. 241-42.)

For the general bibliography and appendices of the First Edition I have substituted at the end of each chapter specific suggestions for further study of the subject under consideration. The enormous increase during the past decade of literature on the subjects discussed in the present book makes it necessary for the student in the future to rely on annotated bibliographies indicating which studies are likely to contain material relevant to a particular inquiry. On the subject of attitude research alone, for example, there appeared in periodicals in English, during the years 1937–39 inclusive, some two hundred titles, not counting relatively inaccessible theses and papers and *excluding* all studies having no bearing on methodology. Annotated bibliographies for the field have, fortunately, become increasingly available, enabling students to go more directly to the relevant sources. A bibliography of such bibliographies, aggregating many thousands of titles, is appended to the present volume.

My first obligation to be acknowledged in connection with the preparation of the present text is to the users of the First Edition. Their criticisms and suggestions are chiefly responsible for the improvements which will be found in the present volume. Among the many who have made helpful suggestions in this way are Raymond F. Sletto of the University

of Minnesota, Paul F. Lazarsfeld of Columbia University, Stuart C. Dodd of American University of Beirut, Ruth Inglis of the Research Division of the Curtis Publishing Company, J. E. Bachelder of the University of New Hampshire, and Calvin F. Schmid of the University of Washington. I am indebted to the first three of these persons and to F. F. Stephan of Cornell University for permission to quote liberally from their published works. Mr. Sletto and Miss Inglis have also read parts of the manuscript and made valuable suggestions. Norman E. Himes, in his capacity as editor of Longmans Sociology Series, has greatly improved the form of the final manuscript by numerous valuable criticisms. I am indebted also to Robert Bierstedt of Bard College, Columbia University, for his painstaking reading of the proofs. My greatest obligation is to Harry Alpert of the College of the City of New York, who has not only read the entire original manuscript, but has given generously of his time in written and oral discussion in the course of which he has made valuable criticisms and suggestions. My indebtedness to all of the above and to others who have shown interest in the revision implies, however, no responsibility on their part for the faults and limitations which remain.

GEORGE A. LUNDBERG

North Bennington, Vermont
June 1, 1941

CONTENTS

CHAPTER PAGE

I. The Theory and Planning of Social Research 1

A. Fields and Methods of Science, 1; B. Levels of Scientific Procedure, 5; C. Steps in the Scientific Method, 9; D. Common Misconceptions Regarding the Nature of Science and of Scientific Law, 11; E. The Possibility of a Science of Sociology, 14; F. The Complexity of Social Data, 16; G. The "Subjectivity" and "Intangibility" of Social Phenomena, 19; H. Predictability of Social Behavior, 21; I. Quantitative Methods in the Social Sciences, 22; J. Quantitative Methods and Case Work, 25; K. Laboriousness of Quantitative Research, 27; L. The Selection of Research Problems, 31; M. Planning a Research Project, 34; N. Summary and Conclusion, 39; O. Suggestions for Further Study, 42

II. Difficulties of Objective Observation. Experimental Techniques 45

A. The Nature of Knowledge, 45; B. The Nature of Bias and Prejudice, 46; C. The Prevalence of Bias and Prejudice in the Social Sciences, 47; D. Inadequacy of Subjective Judgments, 47; E. The "Common-Sense" Fallacy, 48; F. The Development of Objectifying Devices, 49; G. Types of Objectifying Devices, 50; H. Objective Trends in the Social Sciences, 51; I. Intellectual Honesty and the Scientific Method, 52; J. The Non-Ethical Nature of Science, 53; K. The Experimental Method in the Social Sciences, 54; L. The Utopian Community "Experiments," 57; M. Social Legislation as Social Experimentation, 58; N. Scientific Social Experimentation, 60; O. Other Experimentation and Refined Observation, 63; P. Social Experimentation and the Statistical Method, 65; Q. Summary and Conclusions, 75; R. Suggestions for Further Study, 79

CHAPTER PAGE

III. Terminology, Units, and Classification 80

A. The Importance of Definite Terminology in Science, 80; B. Confusion of Terminology in the Social Sciences, 83; C. The Criteria of Scientific Terminology, 86; D. The Nature of Operational Definitions in Science, 88; E. Operational Definitions in the Social Sciences, 91; F. The Units of Science as Constructs of Convenience, 93; G. Requirements of Desirable Units, 96; H. Classification, 104; I. Summary and Conclusion, 109; J. Suggestions for Further Study, 111

IV. The Principal Methods of Social Research 113

A. The Similarity of Different Methods, 113; B. The Historical Method and Natural Science, 114; C. Complementary Nature of Different Methods, 116; D. The Nature and Function of Working Hypotheses, 117; E. Dangers in the Use of Hypotheses, 120; F. Classification of Methods of Collecting Data, 121; G. Documentary Sources, 122; H. Critical Evaluation of Documentary Sources, 125; I. Methods of Field Investigation, 128; J. Summary and Conclusion, 131; K. Suggestions for Further Study, 133

V. The Sample in Social Research . . . 134

A. The General Use of Sampling, 134; B. Conditions Governing the Validity of a Sample, 135; C. Methods of Sampling, 137; D. The Size and Reliability of the Sample, 145; E. Practical Tests of the Adequacy of a Sample, 149; F. General Cautions in Sampling, 149; G. Summary, 153; H. Conclusion, 155; I. Suggestions for Further Study, 157

VI. The Schedule as an Instrument of Observation 159

A. The Need of Instruments of Observation in the Social Sciences, 159; B. The Importance of the Schedule as an Aid to Observation, 160; C. Types of Schedules, 162; D. Scope and Purpose of the Investigation, 162; E. The Construction of Schedules, 165;

CHAPTER PAGE

F. The Testing of Schedules, 175; G. The Filling Out of Schedules, 175; H. Advantages of Personal Investigation, 178; I. Conclusion, 180; J. Suggestions for Further Study, 180

VII. The Questionnaire 182

A. The Questionnaire as a Special Type of Schedule, 182; B. The Solicitation of Cooperation, 183; C. The Content and Structure of the Questionnaire, 186; D. Wording of Questions, 190; E. Physical and Esthetic Aspects, 196; F. Pretesting of Questionnaires, 197; G. Reliability, 198; H. Validity, 201; I. Sending out the Questionnaire, 202; J. Follow-up Letters, 203; K. Who Answers Questionnaires, 205; L. Conclusion, 208; M. Suggestions for Further Study, 209

VIII. The Measurement of Attitudes and Opinions 211

A. Practical Aspects of the Measurement of Attitudes and Opinions, 211; B. Definitions of Attitudes and Opinions, 212; C. The Measurability of Attitudes, 215; D. The Relation of Verbal Attitudes to Other Behavior, 217; E. Methods of Studying Attitudes, 220; F. Attitude Questionnaires and Polls, 221; G. Arbitrary scales, 223; H. Experimental Scales, 234; I. Estimating the Validity of Tests, 242; J. Summary and Conclusion, 246; K. Suggestions for Further Study, 251

IX. The Measurement of Institutional Behavior 253

A. The Demand for Measurements of Institutional Behavior, 253; B. General Requisites of a Sociometric Scale, 254; C. Problems in the Construction of a Sociometric Scale, 258; D. The Factors and Weights in a Sociometric Scale, 260; E. Summary of the Principles of Construction of Sociometric Scales, 265; F. The Measurement of Hygienic Behavior, 270; G. A Scale for the Appraisal of City Health Work, 279; H. The Measurement of Socioeconomic Status, 284; I. The Principal Steps in the Construction of Sociometric Scales, 294; J. Conclusion, 306; K. Suggestions for Further Study, 308

CHAPTER PAGE

X. Sociometric Methods in Ecology and Interpersonal Relations 310

A. Types of Symbolic Representation of Social Phenomena, 310; B. Techniques of Human Ecology, 315; C. Sociographic Methods, 325; D. Sociological Indices, 332; E. Summary and Conclusion, 343; F. Suggestions for Further Study, 345

XI. Field Work: The Interview and the Social Survey 349

A. Introduction, 349; B. The Authority and Backing of the Field Worker, 350; C. The Organization of a Field Staff, 352; D. The Interview, 359; E. The Group Interview, 371; F. The "Panel," 371; G. The Participant Observers, 375; H. The Life History or Case History, 379; I. The Social Survey, 388; J. Conclusion, 392; K. Suggestions for Further Study, 394

XII. Social Bookkeeping 397

A. Social Organization and the Gathering of Social Data, 397; B. The Development of National and International Statistics, 398; C. The United States Census Bureau, 400; D. Trends Toward Increasing Scope and Standardization of Social Statistics, 402; E. Obstacles to Comprehensive and Standardized Social Bookkeeping, 406; F. Social Change and Standardized Statistics, 408; G. Conclusion, 409; H. Suggestions for Further Study, 412

Bibliography of Bibliographies 415

Index 417

TABLES

TABLE		PAGE
I.	Composition and Proportioned Sample of a Student Body	142
II.	Proportion of Replies to Successive Waves of a Questionnaire Regarding Radio Programs	204
III.	Proportion of Replies to Successive Waves of a Questionnaire on Radio Broadcasts	205
IV.	Comparison by Number and Percentages of Attitude Studies Published 1925–1931 and 1937–1939	250
V.	Frequency Distribution of Ratings of Socio-Economic Status of 196 Families by (a) The Chapin Scale (b) a Janitor, and (c) a Banker	292
VI.	Frequency of Types of Relationships in a Group	338
VII.	Some Personal Differences on Which Refusals, Cooperators and Unavailable Groups Differed	362

FIGURES

NO.		PAGE
1	A Simple Type of Schedule—A Birth Registration Schedule	163
2	Population Schedule	*Between pages* 168 and 169
3	Schedule for the Investigation of an Industrial Establishment	170
4	Examples of Desirable Revisions of Questions Appearing on a Church Questionnaire	172
5	Example of a Questionnaire	187
6	Sample Pages from the Appraisal Form for Local Health Work	284
7	The Social Status Scale 1933	290

NO. PAGE

8 Comparative Socioeconomic Rating of 196 Homes by a Janitor, and a Banker and the Chapin Scale 293
9 Rural Cultural Regions and Subregions in the United States 318
10 Rates of Delinquency Cases by Mile Zones Surrounding the Loop (Chicago) 321
11 Juvenile Delinquency Map 323
12 Comanche Social Structure 327
13 Friendship Constellation in a Village . . . 333

SOCIAL RESEARCH

SOCIAL RESEARCH

CHAPTER I

THE THEORY AND PLANNING OF SOCIAL RESEARCH

If there are some subjects on which the results obtained have finally received the unanimous assent of all who have attended to the proof, and others on which mankind have not yet been equally successful; on which the most sagacious minds have occupied themselves from the earliest date, and have never succeeded in establishing any considerable body of truths, so as to be beyond denial or doubt; it is by generalising the methods successfully followed in the former enquiries, and adapting them to the latter, that we may hope to remove this blot on the face of science. — *J. S. Mill.*[1]

A. FIELDS AND METHODS OF SCIENCE

Social scientists are committed to the belief that the problems which confront them are to be solved, if at all, by judicious and systematic observation, verification, classification, and interpretation of social phenomena. This approach in its most rigorous and successful form is broadly designated as the scientific method. The principal techniques and considerations that have been found most useful in the application of this method to social behavior are the chief subjects of this book.

The degree to which the scientific approach to social problems has penetrated our generation (which is frequently called a "scientific age") is perhaps greatly exaggerated. It is true that scientific information about many subjects, particularly about the physical world, is more widely diffused than ever before. It is true that the term "scientific" carries with it a prestige which makes it a much desired and a much misused label. But the *scientific habit of thought,* especially about social problems, has not penetrated very far. We still prefer to deal with these problems in the "good" old manner of our primitive ancestors. That is, we like to ascribe social problems to the malevolence of some persons and the solu-

[1] *A System of Logic,* (8th Edition, Longmans, 1930), p. 546.

tion of difficulties to the benignity of others. Political conflicts, including wars, are thought of almost entirely in this primitive framework.

The gullibility of vast numbers of otherwise apparently intelligent people to the propaganda, quackery, and improbabilities of the day is perhaps the best index of the degree to which scientific habits of thought, as contrasted with a lot of facts *about* science, have penetrated. Thousands of people, including the so-called educated, find it quite reasonable to believe and act upon a radio report of an invasion from Mars.[2] In times of war, again, vast numbers of people find their membership in a "scientific age" no preventive to believing the most fantastic tales regarding the diabolical personal and social characteristics of their enemies, to say nothing of the monstrous propensities of other people's forms of government.

The congeniality of this mode of thinking undoubtedly springs from our long habituation to thinking about human affairs in terms of Devils, Gods, Evil, Sin, and other such concepts. By virtue of its simplicity, as Einstein has pointed out,[3] this theological thoughtway is understandable even by "the most undeveloped mind." For this and other reasons, this mode of thought unquestionably has given through the ages a certain solace to man by imposing a kind of order upon the universe and the promise of the ultimate achievement of our desires. We shall not review here the course of events by which this type of orientation has been superseded gradually as regards the physical world by the scientific type of thinking. As regards the social world, however, our emancipation from the thoughtways of our remote ancestors is quite superficial. Having divested ourselves of a few *individual examples* of our Sunday-school superstitions, many of us, including social scientists, feel very much liberated, sophisticated, and "scientific." The superficiality of our emancipation from the *fundamental thoughtways,* of which the belief in ghosts, hell-fire, etc., are merely *illustrations,* becomes painfully evident, however, in any period of even moderate social crisis.

[2] H. Cantril, *The Invasion from Mars,* Princeton University Press, 1940.
[3] "Science and Religion," *Science News Letter,* Sept. 21, 1940, p. 181.

At such times the agencies of propaganda have little difficulty in convincing even a large number of scientists of the diabolical intentions of "The Enemy" for world conquest, enslavement of the people, and the danger of the ultimate triumph of sin and evil, according to sequences specified in familiar thought patterns. To raise the question as to the *probability* of these eventualities, in view of the known tendencies of individual and social action, is equivalent to siding with "The Enemy." When the crisis is over, to be sure, the absurdity and credulity of our recent behavior is pointed out by students of the subject and is widely accepted by the public. So superficial is the habit of rational and scientific thinking about social events, however, that in situations at all critical, even the scientifically trained promptly abandon this comparatively recent acquisition in favor of the "marvelous articles of savage-born lore" which the conditionings of childhood, and generations of pre-scientific thinking have engrained in us in a way that it will take centuries for science to overcome.

To change these traditional ways of thinking about men and social events, rather than the mere diffusion of information, is undoubtedly the principal task of scientific education. At present, enormous energy reserves in the form of highly undisciplined emotions of hate and love are generated and dissipated through extremely primitive instruments of social organization, consisting fundamentally of an inadequate technology of thought and action regarding social problems. Brotherly love, on the one hand, and hatred of "evil," on the other, are invoked in the face of problems that will yield, if at all, only to instruments of precision like "those of the surgeon and the engineer in hands that do not shake with fear, with anger or with love." The reluctance to discipline these satisfying and virtuous emotions in our dealing with social questions is undoubtedly one reason for the slow permeation of the scientific method into the field of social relations.

Another and more specific reason is the nature of our scientific education and our notion as to what science is. Inquiry even among college students will reveal that science is generally thought of as a *type of subject matter* rather than

as a *method* of study. To some, science means colored liquid in a glass tube; to others, the paraphernalia of the physics laboratory; to still others, it signifies a terminology liberally interspersed with mathematical formulas. The idea of science as a particular method of study, a definite set of rules of procedure and of logic applicable to *any* subject matter has been neglected in our preoccupation with the startling findings and achievements of the better developed sciences. We shall use the term "science" in this book chiefly to designate a method, although we shall occasionally use it, as is common practice, to designate also any body of knowledge arrived at by this method.

The question as to what fields of human knowledge may properly be dignified by the name of science is still largely a matter of controversy and opinion. Some of the older and more highly developed fields of knowledge, such as astronomy, chemistry, and physics, are conceded by practically everyone to be properly classifiable as sciences. Outside of these well charted departments, controversy over the proper application of the term increases in proportion to the difference in subject-matter and the newness of the more recently developed fields. Of course this controversy can never be absolutely settled, because the criteria of science always will be stated in relative terms. For example, it would be impossible to define exactly at what point chemistry or physics passed from magic and alchemy to the status of science.

Fortunately it is not necessary for our purpose to lay down such absolute definitions. There are some general criteria of scientific method which, in spite of the differences of opinion referred to above, are sufficiently well accepted to permit us to know when and to what degree they have been applied in a given field. But when a general criterion or definition of science is attempted it is found that such definition tends to be in terms of *method* rather than of subject-matter. All that the term "science" as applied to a particular field comes to mean is a field which has been studied according to certain principles, *i.e.,* according to *scientific method.* If our knowledge of a certain field has been derived according to this method, and if that knowledge is applicable to this field for purpose of prediction and control, then that body of

knowledge may properly be designated as a science, to the extent that it has been scientifically derived, regardless of the nature of its subject-matter. The test of the thoroughness with which the method has been applied is found in our ability to predict the behavior of classes of phenomena *under given conditions.*

The attempt to define science in terms of subject-matter causes only confusion. The important idea to emphasize is that science is a unity, and that the various sciences are merely convenient differentiations of this unity on the basis of the different problems to which this common method is applied. Says Karl Pearson:

> The scientific method is one and the same in all branches. . . . *The unity of all science consists in its method, not in its material alone.* The man who classifies facts of any kind whatever, who sees their mutual relation and describes their sequences, is applying the scientific method and is a man of science. The facts may belong to the past history of mankind, to the social statistics of our great cities, to the atmosphere of the most distant stars, to the digestive organs of a worm or to the life of a scarcely visible bacillus. It is not the facts themselves which make science, but the method by which they are dealt with.[4]

This is a consideration also, which might be reflected upon with profit by those who are inclined to think of science as a method applicable only to the fields in which it is conspicuously successful at the present time.

Science, then, is fundamentally a technique of deriving reliable knowledge about any type of phenomena in the universe and then applying this derived knowledge for the purposes of prediction and control.

B. LEVELS OF SCIENTIFIC PROCEDURE

We have said that broadly speaking, scientific method consists of systematic observation, classification, and interpretation of data. Now, obviously, this process is one in which nearly all people engage in the course of their daily lives. The main difference between our day to day generalizations and the conclusions usually recognized as scientific, lies in the degree of formality, rigorousness, verifiability, and general

[4] Karl Pearson, *The Grammar of Science,* (Third Edition, A. and C. Black, London, 1911), pp. 10, 12.

validity of the latter. This is, however, necessarily a matter of degree. We may for convenience recognize four degrees or levels of this process:[5]

(*a*) The simplest and crudest method of research is random observation — merely trusting to chance that some valuable datum or suggestion may turn up in the course of miscellaneous exploratory surveys, studies, or experiments. A great deal of social research is still in this stage and, while we should not be content to keep all or most of our research on this level, it is not to be ignored. Even the more developed sciences are always on the alert for curious but significant, unexpected, or unlooked-for phenomena which may suggest hypotheses and more systematic investigations of great importance. The discovery of the x-ray is a classical example of the occasionally enormous significance of accidentally encountered observations.

While recognizing that such random observation is sometimes productive, we must remember that the method is fruitful chiefly in suggesting hypotheses, the verification of which still depends upon the more formal and systematic methods discussed below. We must not be misled into thinking that because a lucky guess sometimes turns out to be scientifically demonstrable, that guessing is therefore a way of arriving at scientific generalizations. Nor must we allow the above recognition of the possible value of random observation to commit us to the cliché that because "all science is a search for truth" therefore (a frequently unspoken *non-sequitur*) all search for truth (by whatever methods) is science.

(*b*) A second method of very wide and useful application in research is systematic exploration of broad fields or subjects. It differs from the first method chiefly and importantly in being *systematic*. This seems to be the method advocated by Francis Bacon in his *Novum Organum*.[6] In modern times, the discovery of salvarsan by Ehrlich perhaps illustrates in a general way this very important type of research. Most social research (*e.g.*, community studies like

[5] Adapted from C. L. Hull, "The Conflicting Psychologies of Learning — A Way Out," *Psych. Rev.*, 42:493-94, 1935.

[6] *Advancement of Learning and Novum Organum,* (Revised Edition, Colonial Press, 1900.)

the Lynds' *Middletown*) is probably now conducted on this level.

(*c*) A third level of scientific research is the testing of well defined but isolated hypotheses, either by experiment or by statistical methods. An increasing proportion of social research is gradually attaining this level of specificity, and a great many of the better established generalizations have been arrived at or are being tested by this procedure. This level of research differs from the preceding chiefly in the degree of *explicitness* with which the hypotheses are stated. All gathering of data involves an *implicit* hypothesis which causes us to select some data while neglecting others. Thus, the taking of a census involves at least the hypothesis that such data will be useful in the future, as they have been in the past. On the other hand, a study of the relationship between the education of a population and its birth rate involves a more explicit and definite hypothesis. Some of the recent studies on housing, mobility, morale, and projects testing the relationship between economic condition and marriage rates are examples of this type of research.[7]

(*d*) The fourth and most advanced level on which scientific research is carried out is found in experiments or other crucial compilations of data, which are directed by systematic and integrated theory rather than by isolated and vagrant hypotheses. This method begins with (1) a set of rigorous and unambiguous terms to be used. (2) Next is stated a set of postulates or hypothetical statements and their corollaries (inferences from the postulates) which are tentatively assumed to be true for the purposes of the investigation. (3) Thirdly, theorems are stated as formal propositions which could or should be true if the postulates, the corollaries and the reasoning in the theorems are sound. These theorems are really hypothetical generalizations, which direct the nature of the (4) empirical observations undertaken to test the validity of the entire explicit theoretical structure outlined above.

A classical example of this procedure is, of course, Isaac

[7] See, for example, S. A. Stouffer, "Intervening Opportunities: A Theory Relating Mobility and Distance," *Amer. Sociol. Rev.*, 5:845-67, 1940; F. S. Chapin, "An Experiment on the Social Effects of Good Housing," *Ibid.*, 868-79; D. C. Miller, "Morale of College Trained Adults," *Ibid.*, 880-89.

Newton's *Principia*.[8] More recent examples are found in the systematic theoretical development of relativity by Einstein,[9] and of quantum theory[10] by Bohr, Rutherford, Heisenberg, Schrödinger, Dirac and others. A well known illustration of investigation motivated by relativity theory is the astronomical observation whereby it was demonstrated that the image of a star, whose light rays had passed close to the sun, showed a certain amount of displacement from its true position, conforming both in direction and amount with deductions made from the theory.[11]

This method sharpens the research problem and the conditions of its solution to a point where the work of different empirical observers can be rigorously evaluated as either corroborative or contradictory. This formal procedure also facilitates the detection of the reason for any discrepancy found by different investigations or between theory and empirical fact.

By this procedure, disagreements between two or more scientists can be resolved by a set of well defined steps rather than by mere argument, recrimination, and quarreling. First of all, it can be determined whether each scientist uses the same definitions and whether each uses them consistently. Secondly, the logical steps can be checked against the accepted rules of reasoning. If the deductions are logically invalid, the theory is unsound, regardless of the facts. Thirdly, if the verification of the deductions involves conditions of observation which are impracticable or impossible of attainment, the theory is metaphysical rather than scientific. Finally, if the deduced result is not corroborated *when the conditions of the theory are fulfilled,* the theory is false. A discrepancy between the deduced results and the observations may, however, be due to a flaw in any stage of this process — postulate, definition, reasoning, or observation. To bring harmony between theory and observation it may be necessary to change the postulates, definitions, and/or rea-

[8] *Mathematical Principles,* (Trans. by F. Cajori, Univ. of California Press, 1934).

[9] A. Haas, *Introduction to Theoretical Physics,* Vol. II, (Trans. from 3rd and 4th editions by T. Verschoyle: Constable & Co., London, 1925), p. 299.

[10] R. Reiche, *The Quantum Theory,* (Trans. by H. S. Hatfield and H. L. Brose: E. P. Dutton & Co., New York), 1930.

[11] A. Haas, *op. cit.,* p. 370.

soning, assuming now that the observation is reliable. Disagreement between different theories can also be identified and explained in this way.

It is true that this method works best in the developed sciences where rigorous experiment and mathematical statement are possible. Clark Hull[12] has, however, done notable work by this method in psychology, and, as a goal toward which to work, this remains the ideal of scientific procedure in the social as well as in the physical sciences.

C. STEPS IN THE SCIENTIFIC METHOD

We have emphasized above that inquiry which yields knowledge of the type we recognize as scientific may be, and is, carried out on different levels of formality, efficiency, and perfection. On all of these levels, however, there is discernible, explicitly or implicitly, a certain series of recognizable and verifiable steps which taken together constitute the scientific method. They may be summarized as follows:

1. *The working hypothesis.* An hypothesis is a tentative generalization, the validity of which remains to be tested. In its most elementary stages, the hypothesis may be any hunch, guess, imaginative idea or intuition whatsoever which becomes the basis for action or investigation. Even the most detached dream or intuition has its basis in human experience of some kind. We must not be deluded into thinking that, because of the subjective (*i.e.,* private, unverified) nature of these experiences, they are somehow generated spontaneously in the "mind." In searching for fruitful hypotheses, we may draw upon the whole realm of poetry, literature, philosophy, and the large descriptive literature of sociology and ethnology, including the speculative theories of artists and penetrating thinkers who have devoted themselves to a deep, if informal and unsubstantiated, study of man's social relationships. But a *working* hypothesis must be formulated as explicitly and definitely as possible. It is in its elementary form merely a definition of a problem or the object of any investigation, including a formulation of what facts

[12] C. L. Hull and others, *"Mathematico-Deductive Theory of Rote Learning,"* (Yale University Press, 1940). An outline of a project proposing to follow this procedure in anthropology will be found in G. P. Murdock, "The Cross-Cultural Survey," *Amer. Sociol. Rev.,* 5:361-70, 1940.

will be necessary to test the hypothesis. As such, it limits the study to proportions that can be handled with the resources at hand and thus prevents aimless and random gathering of data. Herein, of course, lies a grave danger. If the investigator conceives of his hypothesis as a proposition to be defended, to the exclusion of evidence that might support the opposite, or some other hypothesis, this, of course, vitiates his work as unscientific. This danger, and the means of avoiding it, will be further discussed in a later chapter.

2. *The observation and recording of data.* This step includes the definition of units of enumeration and measurement. It consists of making numerous and accurate observations with the human senses (*i.e.,* responding to certain stimuli with frequency and consistency), assisted and corrected by instruments of precision. The technique of achieving accuracy and objectivity in these observations is the main subject of this book. Upon the accuracy of the original observations depends the value of the refined analysis and the ultimate reliability of generalizations drawn from the data.

3. *The classification and organization of the data collected.* It is only when facts are classified that uniformities, sequences and differences become apparent. These uniformities and sequences are the center of interest to science. Whenever uniformities and sequences can be demonstrated in any field, that field at once begins to come within the domain of science.

4. *Generalization.* When data have been collected and classified so as to reveal the patterns, uniformities, and sequences, there remains the task of formulating a brief statement or *description* of these sequences in such form as to apply to all similar phenomena in the universe studied *under given conditions.* This is a scientific "law." The ideal form of such a generalization is the mathematical formula. This formula is never absolute or exact. Even in the fields where scientific investigation has long been employed, the formulas and "laws" are subject to constant refinement and correction on the basis of more extensive and more accurate observation of the data. These corrections usually take place as a result of experimental testing of our generalizations. But the fact

that our formulas may not be exact does not, as we shall see, vitiate the value of such generalizations as a basis of scientific prediction, since the probable error is itself calculable. Caution must be exercised, however, in not applying generalizations too broadly, *i.e.*, not applying them to situations which differ in important respects from those for which the generalizations have been found valid. When this caution is observed, generalizations in turn become the basis for scientific theories of phenomena regarding which data or sufficient data are, as yet, wanting.

D. COMMON MISCONCEPTIONS REGARDING THE NATURE OF SCIENCE AND OF SCIENTIFIC LAW[13]

Although the steps outlined above are reasonably definite and concrete, there are certain widespread misconceptions regarding them which should be noted. One of these misconceptions is that a scientific law is a so-called "natural law" inherent in the universe, perhaps laid down by some supernatural agency, and waiting to be "discovered" by some bright person. Thus Newton is credited with "discovering" the law of gravity. What he did was to describe certain uniformities in the behavior of falling bodies. *This generalized description of the behavior of phenomena under given conditions* constitutes the scientific law, not the sequences, forces, or behavior of phenomena themselves, which, of course, existed before Newton stated his law. In other words, a scientific law does not exist until it is stated, although the *behavior* and relationships which it describes may have existed for millions of years. Says Pearson in a brilliant passage:

> A scientific law . . . is the *résumé* or *brief expression* of the relationships and sequences of certain groups of . . . perceptions and conceptions, and exists only when formulated by man. . .
>
> We are thus to understand by a law in science, *i.e.*, by a "law of nature," a *résumé* in mental shorthand, which replaces for us a lengthy description of the sequences among our sense-impressions. Law in the scientific sense is thus essentially a product of the human mind and has no meaning apart from men. . .
>
> The reason we find in natural phenomena is surely put there

[13] For a more complete and adequate development of this subject, see my *Foundations of Sociology*, ch. 4.

by the only reason of which we have any experience, namely the human reason . . . the logic man finds in the universe is but the reflection of his own reasoning faculty.[14]

A further confusion as to the meaning of law in science results from applying the term to propositions of pure theory or logical truisms as well as to generalizations from empirical observations. Thus, such a proposition as the "law" of inertia, to the effect that matter will remain at rest unless some force acts upon it, is obviously true if we define force as *that which produces motion.* Likewise, Spencer's "law" that social motion follows the line of least resistance is obviously true if "the line of least resistance" is always defined as that line which the motion actually did take as contrasted with other possible lines it might have taken. Such scientific tautologies, truisms or axioms may be valuable as hypotheses or postulates in giving framework and direction to investigation. They should not be confused, however, with empirical laws derived from the statistical generalization of empirical observations, describing how phenomena will actually behave under stated conditions. As Hutchison[15] has said:

> The prevailing tendency to call propositions of pure theory "laws" is misleading and inappropriate, and appears to be a survival from eighteenth-century rationalist philosophy, and theology. . . We suggest that the term "law" should be reserved only for those empirical generalizations such as Pareto's or Gresham's law or the law of diminishing returns, or diminishing marginal utility. It is such laws as these that it is the central object of science to discover.

A second serious misconception of the meaning of scientific law is that which holds it to be an exact statement of uncontrolled behavior *under all conditions.* As a matter of fact, scientific laws are nothing of the sort. They merely state how phenomena behave *under stated conditions.* These conditions are sometimes highly abstract and artificial, and do not obtain anywhere in the uncontrolled universe. Thus, to use Bernard's illustration, the law of falling bodies de-

[14] Karl Pearson, *op. cit.*, pp. 82, 86, 87, 91.
[15] T. W. Hutchison, *The Significance and Basic Postulates of Economic Theory,* (Macmillan, 1938), pp. 63, 64.

scribes how bodies fall *in a vacuum, at sea-level.*[16] In other words, only under artificial conditions does this law hold good as a description of the behavior of falling bodies. The fact that these conditions nowhere obtain in the natural universe does not in any way vitiate the value of this law for scientific purposes. *It furnishes a norm or a principle from which variations may be computed or estimated.*

A third misconception regarding the nature of science and scientific law is the idea that it is the function of science to "explain" behavior, in the sense of accounting for it in terms of first causes. This is, of course, a pretension which no scientist has ever advanced, and which no science has ever accomplished. All that science pretends to do is to describe certain sequences in the behavior of the phenomena it studies. In the physical sciences, a description of these sequences is usually accepted as constituting the only kind of explanation we know anything about. But in the social studies, we find learned discussions of the limited usefulness of the methods of natural science because all science can do is to describe, whereas in the social sciences we must also have *explanation* in the sense of some *ultimate* cause. This is a survival of a pre-scientific mode of thought. The only explanation known to science is conceptualized description.

As a matter of fact that is what all explanation consists of. We say that anything is "explained" or "understood" when we have reduced a situation to elements and correlations with which we are so familiar that we accept them as a matter of course so that our curiosity rests. By "element" we mean any component which we do not consider it necessary or possible further to analyze. Understanding a situation means, from the operational point of view, discovering familiar elements, and correlations between them.[17]

With reference to more "ultimate" kinds of explanation, Veblen's remarks are pertinent:

[16] L. L. Bernard, "The Objective Viewpoint in Sociology," *Amer. Jour. Sociol.*, 25:307, 1919. This is a very frequently overlooked point. See W. M. Kozlowski, "The Logic of Sociology," *Amer. Jour. Sociol.*, 33:912-21, 1928. Also Eleanor Bisbee "Objectivity in the Social Sciences," *Phil. of Sci.*, 4:371-82, 1937.

[17] *Cf.* P. W. Bridgman, *The Logic of Modern Physics*, (Macmillan, 1932), p. 37.

Questions of a primordial beginning and a definite outcome have fallen into abeyance within the modern sciences, and such questions are in a fair way to lose all claim to consideration at the hands of the scientists. Modern science is ceasing to occupy itself with the natural laws — the codified rules of the game of causation — and is concerning itself wholly with what has taken place and what is taking place. . . The sciences which are in any peculiar sense modern take as an [unavowed] postulate the fact of consecutive change. Their inquiry always centers upon some manner of process . . . and comes to rest provisionally when it has disposed of its facts in terms of process. But modern scientific inquiry in any case comes to rest only provisionally; because its prime postulate is that of consecutive change and consecutive change can, of course, not come to rest except provisionally. By its own nature the inquiry cannot reach a final term in any direction. So it is something of a homiletical commonplace to say that the outcome of any serious research can only be to make two questions grow where one question grew before.[18]

E. THE POSSIBILITY OF A SCIENCE OF SOCIOLOGY

The steps and considerations outlined above are generally recognized as the essence of scientific method. The validity of this method for the understanding and control of physical phenomena is generally conceded. If one casts even a superficial glance over modern western civilization there is one fact which stands out above all others in significance, namely, the astonishing development of physical science and its employment in the control of the physical world.

The same glance reveals the relatively chaotic and uncontrolled state of affairs in the field of human relations. Every day there is a new tale of conflict, bloodshed, disorder, crime and misery, owing to man's imperfect control of his relations with his fellowmen. It is only very recently that the idea has gained currency that perhaps the method which has been found so successful in bringing the physical world under man's understanding and control might also be a profitable one to pursue toward the understanding and control of social relationships.

With the development of this idea, the question as to the applicability of this method to social phenomena has become of increasing interest. In the controversy and speculation on

[18] Thorstein Veblen, *The Place of Science in Modern Civilization,* (Viking Press, 1932), pp. 32, 33, 37, 38.

this point, a number of objections are presented by those who take the negative view, which seem to them conclusive evidence of the impossibility of a social science in anything like the sense in which the term is employed in chemistry or physics. Therefore, a careful consideration of these objections is desirable at this point.

Most of the objections against the idea of exact social science or sciences flow from the feeling that there are certain inherent differences in physical and social data precluding or largely invalidating the employment of the scientific method in the study of social phenomena. This objection, as generally held, reflects the deep-seated notion, already referred to, that science is, after all, a body of subject matter rather than a method. The more specific and detailed objections are to a certain extent corollaries of this central issue. It will be well to consider further, therefore, the differences between social and physical phenomena which are alleged to be fatal to the formulation of scientific laws of social behavior.

To begin with, the history of science furnishes some highly suggestive and illuminating data on the question of the applicability of scientific methods to new fields. The question as to what is the proper or possible field of science is not a new one. It is as old as science itself. It appears that the proper and possible field of science has always been defined so as to coincide strictly with the domain within which science has, at the time, established itself and secured recognition. Says Pearson:

> Wherever science has succeeded in ascertaining the truth, there, according to the school we have referred to, are "the legitimate problems of science." Wherever science is yet ignorant there, we are told, its method is inapplicable; there some other relation than cause and effect (than the same sequence recurring with the like grouping of phenomena), some new but undefined relationship rules. In these fields, we are told, problems become philosophical and can only be treated by the method of philosophy. . . Now it will, I think, be found that the fields of inquiry, where science has not yet penetrated and where the scientist still confesses ignorance, are very like the alchemy, astrology, and witchcraft of the Middle Ages. Either they involve facts which are in themselves unreal — conceptions which are self-contradictory and

absurd, and therefore incapable of analysis by the scientific or any other method — or, on the other hand, our ignorance arises from an inadequate classification and a neglect of scientific method.[19]

So much for the general barring of scientific method from certain fields on the ground of fundamental differences in the phenomena.[20] But there remain some more specific objections based on apparent difficulties in the application of scientific method of social phenomena.

F. THE COMPLEXITY OF SOCIAL DATA[21]

Perhaps the most frequently urged obstacle to a true science of human group behavior is the complexity of its subject matter. Human groups in their tremendous sensitivity to physical and social stimuli and in their enormous cultural, psychological, and temperamental attributes which condition their behavior, appear to be such an inextricable combination or confusion that the mind of man is staggered in its attempt to read order, sequence, and law into the behavior of a social group. Stuart Chase has imagined a man's view of North America from an airplane in these words:

> He would only see the farms, the forests, the mines, the railroads and highways, the rivers, canals, transmission lines; the factories, warehouses, stores, schools, libraries, theatres, golf courses, and homes; the behavior of some 100,000,000 men, women and children in relation to these things. Men digging, and plowing, pulling the throttles of engines, balancing on steel girders, painting signboards, holding steel under a drill, wrapping up packages, driving trucks, bending over desks, talking through telephones, jamming people into elevated train doors, fishing on the high seas, fighting forest fires, pumping oil wells, reading newspapers, yelling at ball games, sleeping, eating, lovemaking, going to church, dancing, swimming, climbing mountains — pacing in a striped suit through prison corridors. Women minding spindles and babies and cook stoves, playing Mah Jong, drinking tea, smoking cigarettes. Children answering the school

[19] Karl Pearson, *op. cit.*, pp. 18, 22, 23.

[20] For an elaboration of this position, see my *Foundations of Sociology*, pp. 9-22.

[21] An excellent elaboration of this subject may be found in two articles by Read Bain, "The Concept of Complexity in Sociology," *Social Forces*, 8:222-31, 1929, and 369-78, 1930.

gong, twisting in their seats, rushing through the playground, working in cotton mills, tossing with fever. . . A vast conglomeration of human activity.[22]

Does it not seem unreasonable to hope that the mind of man can ever reduce this chaos to order, and define this extraordinarily complex surging and milling of the human ant-heap into fairly exact scientific laws? On what grounds may such optimism be entertained?

In the first place, there is the fact that in all its complexity there is nevertheless to be observed, even by the casual student, certain patterns, sequences, and order in the multitudinous activities of a social group. Its behavior is by no means haphazard, random, and impossible to predict. On the contrary, if we take a group of which we are a part or which we know well, we find that most of its behavior is of a highly routine, uniform and predictable nature. Given certain conditions of a situation, we find ourselves able to predict with a high degree of accuracy what a group will do. We find that it tends to do certain things at certain times, *e.g.*, eating, sleeping, going to church. It tends to wear certain kinds of clothes on certain occasions. As a matter of fact, our own activities and social organization in general is founded upon this tendency of a social group to behave in a definite manner *under given conditions.*

We find ourselves, therefore, making very accurate estimates and predictions as to the future behavior of human beings and of social groups. Thus, the behavior observed from the airplane, while it may seem complex, is not chaotic after we have studied it a little. It is chaotic, disordered, "meaningless" only to the superficial observer. The more we study it, the more orderly and predictable this behavior becomes. What is a scientific law, but a brief description of

[22] Stuart Chase, *The Tragedy of Waste,* (Macmillan, 1930), pp. 17, 18. It is pertinent to know that the physical scientists are equally impressed with the complexity of their subject matter. Thus, Sir Ernest Rutherford in his presidential address to the British Association for the Advancement of Science (1923) said: "When we consider the extraordinary complexity of the electronic system we may be surprised that it has been possible to find any order in the apparent medley of motions." Quoted in H. Buckley, *A Short History of Physics,* (Van Nostrand, 1928), p. 254.

how phenomena behave under given conditions? This is not only possible but has been practiced in a more or less systematic way from the very beginning of society. All that social science aims to do is to change this practice from a rule-of-thumb procedure to an objective and more exact practice permitting corroboration.

The fact that the hopeless complexity and chaos of social behavior tends to disappear with closer study perhaps suggests the real nature of the complexity generally urged as an insuperable obstacle to an exact social science. *Any situation or behavior of phenomena is complex when we do not understand it.* Complexity is always relative to our understanding or mastery of a given behavior. To the baby who has not yet learned to walk or speak, the perambulations and conversation of its elders constitute behavior of incredible complexity. Complexity is not definable in absolute or objective terms. It is a subjective measure of the degree of adjustment to a situation. *The complexity of human society, in other words, is largely a function of our ignorance of it.*

It is true that the number of variables involved in some situations is greater than in others, and that complexity may be meaningfully defined in these terms. But the definition of a situation is itself an act of discrimination with reference to the adjustment techniques we possess. In short, the complexities of nature have been approached both by simplifying the situations and by developing techniques for handling increasingly large numbers of variables. The maturity of a science is measured by the degree to which it has succeeded in either or both of these respects. If some problems of physics today have fewer variables than some problems of sociology it is because physicists have found it advisable to exercise modesty in the scope of the problems which they set for themselves. They have been content to work patiently on a number of small but related problems. But their efforts in the aggregate have produced the magnificent synthesis of physical science as a whole. Social scientists tend to be impatient of painstaking study of limited problems, and incline rather toward the grandiose methods of the philosophy of history.

G. THE "SUBJECTIVITY" AND "INTANGIBILITY" OF SOCIAL PHENOMENA

Perhaps the most deepseated and widely held reason for insisting on the fundamental difference between social and physical data is the idea that physical phenomena may be known directly through the "senses," whereas many important social phenomena are known only symbolically through words representing such phenomena as tradition, custom, attitudes, values and the whole realm of the so-called subjective world. If this idea were true, it would indeed make an important difference in the two classes of data, for it would mean that we could have much more direct and objective knowledge of physical data because of their intrinsic qualities. But psychological investigation does not bear out the idea that our *knowledge* (in the sense of the power of manipulating symbols which represent sense experience) of physical data is secured in any way essentially different from the way we gain our knowledge of social data. We can have no knowledge of either except through the operation of the symbolic behavior mechanisms, usually the language mechanisms.

The important thing to remember is that such words as tradition, custom, thought, feeling, etc., *stand for observable human behavior* of some kind, and are therefore subject to scientific study by the same general methods as other behavior. It is true that different techniques and instruments have to be used for the observation of different types of behavior. Some behavior we observe directly with the naked eye and report in folk terms. Other behavior is observable only through instruments, for example, microscopes and telescopes. Still other types of behavior are observed through listening through stethoscopes, reading thermometers, or by reading the records of lie detectors, of machines that record "brain-waves," etc. These readings are reported, communicated, and corroborated by means of special words invented for the purpose, such as degrees of temperature, etc.

This point of view also resolves the old problem of introspection and the question as to whether observations of oneself can be used as scientific data. No one questions the

scientific validity of a physician's readings of his own temperature any more than we question his reading of other people's temperatures. His observations are acceptable in either case only if he communicates them to us in a language which permits us to check his findings. The same rule would, of course, hold for his observations of his most intimate thoughts and feelings. When we develop techniques of observing and reporting "subjective" phenomena so that observations can be communicated and corroborated, they are as proper objects of scientific study as any other data. "Objective" and "subjective," therefore, turn out to be not intrinsic qualities of different types of phenomena, but rather a designation of the degree to which we have developed checkable means of reporting our observations.[23]

Thus it appears that the differences between the data of sociology and of the physical sciences are more apparent than real. We do not know physical data any more directly or objectively than social data, except in so far as we have developed more adequate instruments and symbols for recording the former. The uniformities in physical data which form the basis of all physical sciences are also observable in social data. As we have noted, scientific laws specify certain simple and frequently artificial conditions under which phenomena behave in a certain way. Under these conditions, the behavior can be predicted with a high degree of accuracy. Likewise we can predict, with a high degree of accuracy, how many people in a given city will be born, will die, commit suicide, or get married during the coming year, *provided,* the significant conditions obtaining during the past years on which our observations are based remain the same. These conditions must always be specified in our prediction, and might be numerous and involved. Long and patient study of the environment will be necessary before we can specify all the significant conditions. But such reservations are the *sine qua non* of all scientific laws. *The difficulties, therefore, which appear to preclude the possibility of a true science of society, derive from our undeveloped technique*

[23] For a brief elaboration of this subject, see my *Foundations of Sociology,* pp. 17-22. References to a large literature on the subject will also be found in this source.

and methodology of study and our consequent unfamiliarity with the data, rather than from inherent differences in the data themselves.

H. PREDICTABILITY OF SOCIAL BEHAVIOR

A second common obstacle frequently urged against the possibility of the development of a true science of group behavior is the apparent unpredictability of human behavior owing to our time-honored conception of the person as a free agent. Social phenomena are alleged to be subject to unpredictable "whims," due to the "wills" and "minds" of human beings. Fortunately, the behavioristic approach in psychology is rapidly rendering obsolete these quaint metaphysical terms in their historic meaning. The behavior of human beings is recognized as being caused and conditioned in the same general way as the behavior of other phenomena. If from the top of a tower we drop a feather, it falls in a very irregular and unpredictable way. The way in which it falls and the rate at which it falls are determined by a large number of factors, some of the more obvious ones being wind, moisture, and the size and shape of the feather. Still we have not hesitated to apply to it the law of falling bodies, which, as indicated before, describes the falling of bodies only under certain artificial conditions. The unpredictable gyrations, eddies, and spirals with which the feather falls might be attributed to "whims" and "will" with as much (or as little) justification as the behavior of social groups is attributed to such causes. The observed behavior is in both cases the result of environmental factors playing upon a particular type of object. In both cases the environment and the object may be so simplified, artificialized, and standardized as to admit of a brief description of the behavior of the object under these conditions. This is all that the requirements of a scientific law demand.

The apparent unpredictability of group behavior, therefore, is due to our present limited knowledge of the nature of the stimuli and responses operative in such groups. Some of the simpler chemical reactions were considered unpredictable some centuries ago for the same reason. Painstaking observation and classification of these reactions under given

conditions, however, yielded our present chemical laws. The same method has already yielded us some power of generalization and prediction of group behavior. There is every reason to believe that we can vastly extend this power.

We have already referred (p. 20) to the fact that our everyday adjustments to our fellowmen are carried out on the basis of highly reliable predictions of what other people will do. The task of science is to develop more formal and generally reliable methods of such prediction. Encouraging progress has already been made in this direction. Actuarial methods of determining the probability of death, marriage, twin or triple births, divorce, etc., for different age groups, economic classes, etc., have been familiar for some time. More recently the same general approach has been successfully adopted in predicting success or failure in college, in marriage, and in probation and parole.[24] The main problem here as in all scientific prediction is to secure *reliable and sufficiently extensive records of past experience under given conditions.* The best if not the only way to secure such data is through systematic and comprehensive registration of these events as they occur. Only in this way can we modify our predictions according to changing experience. Data on other types of social behavior which do not lend themselves to systematic registration must, of course, be gathered by whatever methods are found feasible.[25]

I. QUANTITATIVE METHODS IN THE SOCIAL SCIENCES

Finally, there is the contention that exact science tends to become increasingly quantitative in its units, measures, and terminology, while most of the important subject-matter of the social sciences is qualitative, and does not admit of quantitative statement. The implied assumption in this position is, of course, that some kinds of data are intrinsically qualitative, subjective, immeasurable.

The untenability of such an assumption in the scientific universe of discourse must be evident from even a slight familiarity with the history and methods of science. Such

[24] See Suggestions for Further Study at the end of this chapter for references to this material.

[25] For elaboration of this subject, see my *Foundations of Sociology*, pp. 142-44.

words as quantitative, measurable, objective, complex, homogeneous, etc., must be regarded, under the postulates on which science proceeds, not as inherent characteristics of data but as designations of certain ways of responding and communicating responses.[26]

The assumed conflict between qualitative and quantitative methods must, then, be abandoned in favor of the view that these terms merely represent different stages of refinement and objectivity in our technique of description.[27] When our knowledge is of a general, unanalyzed kind, we tend to describe it in subjective and qualitative terms on the basis of the feelings which our perception of the facts evokes in us, or perhaps in terms of social evaluations according to existing norms. Thus things may be described as hot or cold, red or green. But these terms mean different things to different individuals; and, as we have seen, one of the essentials of science is that its descriptions shall mean the same to all normally constituted minds. To eliminate the indefiniteness of subjective terminology, the general tendency of science has been to reduce the qualitative gradations to quantitative gradations. Thus hot or cold are described in terms of a thermometer scale; color gradations in terms of light waves; and so on for all the qualities which have become relevant or important in a problem. But in the early stages of our thinking about a field, our descriptions will tend to be qualitative. As Mitchell has pointed out:

". . . Even in the work of the most statistically minded, qualitative analysis will keep a place. Always our thinking will cover a field larger than our measurements; the preconceptions that shape our ends, our first glimpses of new problems, our widest generalizations, will remain qualitative in form. Indeed qualitative work will itself gain in power, scope and interest as we make use of wider, more accurate and more reliable measurements. . .

[26] *Ibid.*, pp. 137-42.

[27] *Cf.* Hutchison, *op. cit.*, pp. 68-69, "Leaving out of account, then, the kind of prognoses given by propositions of pure theory, it does not seem possible to give the terms 'qualitative' and 'quantitative' as applies to prognoses any more sense than 'less accurate' or 'more accurate,' and we suggest that their replacement by these terms would be less confusing."

Cf. also S. C. Dodd, *Dimensions of Society*, (Macmillan, 1941), ch. 1, sec. 4, for a penetrating analysis of the nature of quantitative as contrasted with qualitative symbols and an exact boundary between the two. See also Dodd's ch. 3, sec. 4B.

In the thinking of competent workers, the two types of analysis will cooperate with and complement each other as peacefully in economics as they do in chemistry." [28]

But this concession as to the place of qualitative analysis must not be interpreted as meaning that the distinction between them is unimportant. On the contrary, it emphasizes the view that, for the more exact descriptions required by science, the quantitative statement is necessary. Herein lies the significance of this discussion for the social sciences; for up to the present time they have consisted largely of qualitative generalizations. The explanations and descriptions consist of lists of "desires," "wishes," "wants," "drives," etc. Professor Mitchell's prophecy regarding economics would apply with even greater force to sociology: "If my forecast is valid," he says, "our whole apparatus of reasoning on the basis of utilities and disutilities, or motives, or choices, in the individual economy, will drop out of sight in the work of the quantitative analysts, going the way of the static state. The 'psychological' element in the work of these men will consist mainly of the objective analysis of the economic behavior of groups. Motives will not be disregarded, but they will be treated as problems requiring study instead of being taken for granted as constituting explanations." [29] Only when the quantitative stage is reached do our generalizations begin to partake of the nature of advanced or exact science.

The qualitative worker, therefore, is not open to criticism for the mere fact of being qualitative in his approach, especially in the newer fields. Work of the utmost value may be performed through this approach. But he is open to criticism when he contends that this is the only method applicable to his subject matter, and that therefore his qualitative descriptions represent the only description possible. He may be entirely sincere in his skepticism regarding the possibility of statistical restatement of his generalizations, and in the form in which he has stated them it may indeed be entirely possible that at present they are not susceptible to statistical attack. Says Mitchell: "Indeed . . . there is slight prospect

[28] Wesley C. Mitchell, "Quantitative Analysis in Economic Theory," *Amer. Econ. Rev.*, 15:12, 1925.
[29] *Ibid.*, p. 5.

that quantitative analysis will ever be able to solve the problems which qualitative analysis has framed, in their present form. What we must expect is a recasting of the old problems into new forms amenable to statistical attack. In the course of this reformulation of its problems, economic theory will change not merely its complexion, but also its content." [30]

What will be the nature of this change of content in sociology? As stated above, the present content consists largely of (1) a list of "desires," "wishes," "forces," "interests," or "drives" under the urge of which social groups are alleged to engage in their varied activities; and (2) a bold forcing of the data into these categories on the basis of observed social results of the activities. The difficulty, of course, as Mitchell has pointed out, is that "these interpretations are something which the theorist adds to the data, not something which he draws out of them." As a result, a given behavior is frequently classifiable into any or all of the given categories, according to the prejudice of the classifier. The new content will consist of the statistically summarized results of an analysis of a large number of behavior records in all the fields of social behavior in which we happen to be interested.

J. QUANTITATIVE METHODS AND CASE WORK

But how are such generalizations to be useful in individual cases? This is the fundamental objection of many practical social workers. "No two cases are alike," objects the social worker; "the technique to be employed in different cases is tremendously variable. It varies with every detail of my personality and that of my client. It has to be momentarily adjusted on the basis of actual developments in the course of the interview. How can statistical studies of large numbers of cases, resulting in some 'average' or type, be of value in such practical situations?"

Objections of this kind can be regarded as resulting only from a misunderstanding of what scientific description and statistical technique imply. The facts as stated are, of course, true. But wherein are such practical problems of diagnosis and treatment in any way different from the problems of applied physical science? Is not the physician confronted with

[30] *Ibid.*, p. 3.

the same situation? *Of course,* no two cases are exactly alike in all respects. To be sure, treatment must be modified on the basis of observed reactions during its course. But who would contend that the principles of diagnosis and treatment on which the physician works are therefore useless? And how are these principles developed? They are arrived at by the careful recording of large numbers of different cases and their responses under treatment—a statistical generalization.

As a matter of fact, whether they know it or not, all case workers and technicians do employ an elementary or crude form of statistical method in arriving at the generalizations from which they undertake to make diagnoses and carry out treatment. They are sometimes quite subconscious generalizations based on the technician's memory of the similarities in other cases with which he has had experience. Consider, for example, this statement by an able scholar: ". . . It seems to me that the argument for prediction is bedeviled by a failure to recognize two kinds of prediction: (*a*) actuarial prediction concerning whole populations of people and (*b*) individual prediction concerning the possibilities of action for a single person. . . The displacement of the psychological point of view is the chief fallacy of statistical social science. . . . The only way in which we can predict the chances [*sic*] that a given individual has of behaving in a certain way is to study him as an individual and especially his subjective mental processes with the aid of subjective categories."[31]

That actuarial predictions do not tell *which individuals* will be involved in the predicted behavior is generally recognized by scientists. It does not follow that the *chances* of a given individual's behaving a certain way can be determined by other than actuarial methods. Study, if you will, a given individual as thoroughly as can be imagined, including his most subjective or subconscious mental processes, with the aid of as many subjective categories as anyone desires: What possible basis for prediction does all this material provide *except a basis for classifying the case as more or*

[31] G. W. Allport in *Critiques of Research in the Social Sciences I,* "An Appraisal of Thomas' and Znaniecki's *The Polish Peasant in Europe and America,*" (Social Science Research Council, New York, 1939), p. 185.

less like other cases with which the analyst has had experience (directly or indirectly thru reading about cases, laws of mechanics, physiology, psychology, or what not), all of which represent a formal or informal actuarial basis of *every* prediction regarding "the chances that a given individual has of behaving in a certain way"? What possible basis of prediction could the most intimate knowledge of a case provide, unless the predicter could interpret this knowledge in terms of knowledge of other cases *of behavior*?[32] The value of this intimate knowledge of the case is not denied. Indeed, such knowledge is a necessary basis for refined statistical classification and interpretation. We merely point out that this knowledge gives predictive power only because of the refined classification it makes possible in relation to other cases for which more or less reliable probabilities have already been formally or informally worked out. Accordingly, this objection to quantitative methods from practical workers reduces itself to the question whether the generalizations from which we must proceed should be arrived at by rule-of-thumb methods, or through the definitely quantitative and objective procedure of statistical methods. With the proposition so stated, even the most vigorous critic of the statistical method will scarcely have the temerity to argue. The larger implications of this point for social case work will be treated in a subsequent chapter (Chapter XI).

K. LABORIOUSNESS OF QUANTITATIVE RESEARCH

While it may be conceded that most of the objections which qualitative workers urge against the quantitative approach to social problems flow from a sincere conviction that a statistical approach is impossible, some of the objections derive from less defensible motives. Certainly one of the chief objections to quantitative studies in some quarters is the heavy burden of routine labor usually involved. As Mitchell has pointed out, "a qualitative worker requires

[32] Possibly one reason for the arguments on this point may be a confusion as to what is the unit of observation or "case" in any given instance. The recurrence of a tic, (or other spasm), for example, may be highly individual to some person, and can be predicted from a study of *a large number of occurrences* of the tic *in this person*. But the unit of observation or "case" is clearly the tic rather than the individual, and prediction follows the usual actuarial methods.

hardly any equipment beyond a few books and hardly any helper except a typist. A quantitative worker needs often a statistical laboratory, a corps of computers, and sometimes a staff of field workers." [33] This is not to say that quantitative studies possess any virtue *because* of their laboriousness. Indeed, more elaborate quantitative studies than are necessary to secure results adequate for the purpose at hand are merely a pretentious kind of inefficiency and waste. On the other hand, we should beware of the facile criticism that painstaking quantitative studies show only what "everyone" (or at least the wise) already "knows." Statistical studies do frequently show only what was generally *suspected,* but call attention to the *degree* to which the suspicion was correct and *under what conditions* the facts are as "everyone knows." The distinction is an important one, because social scientists frequently fail to distinguish between what they *know* (as science defines knowledge, *i.e.,* demonstrable fact) and what they "know" as a matter of private conviction, belief and feeling. We may profitably recall the words of Pearson in this connection:

> "There is a certain school of thought which finds the laborious process by which science reaches the truth too irksome; the temperament of this school is such that it demands a short and easy cut to knowledge, where knowledge can only be gained, if at all, by the long and patient toiling of many groups of workers perhaps through several centuries." [34]

The laboriousness of scientific research will be fully appreciated only by those who have experienced it. Some idea of what it involves in the social sciences will be secured from the chapters that follow. Only when an actual project is carried .through, however, will the "dirty work" and the "headaches" of actual research be fully appreciated. Such dismal days and months of routine labor require courage, perseverance, and unflagging devotion to scientific ideals. The satisfaction of "seeing how things come out" is also great; and there are perhaps few students who have not felt this reward in some degree, whether in the solution of a problem in algebra or in some other field.

[33] Wesley C. Mitchell, *op. cit.,* p. 2.
[34] Karl Pearson, *op. cit.,* pp. 17, 18.

The transition to quantitative methods in the social sciences is likely to be deplored by those who feel that in this field, as contrasted with the other sciences, only "literary" and philosophical modes of expression can adequately portray the subtle "human" characteristics of what is called "the situation as a whole." Many of the subjects with which the social sciences deal have been traditionally the province of the preacher, the reformer, the journalist, the novelist, and the dialectician. These facile practitioners of the art of verbalization have sometimes nothing but ridicule for sober scientific work in the social sciences. As the line dividing social science from literature becomes increasingly clear, the difficulty will of course resolve itself in this as in other fields. One of the chief complaints of the literary fraternity is that as the sciences advance, they tend to rely on increasingly specialized symbols and modes of expression which do not lend themselves to oratorical periods and flashy argumentation. In the meantime, Mitchell's prediction for economics undoubtedly also applies to the other social sciences:

> The literature which the quantitative workers are due to produce will be characterized not by general treatises, but by numberless papers and monographs. Knowledge will grow by accretion as it grows in the natural sciences, rather than by the excogitation of new systems. Books will pass out of date more rapidly. The history of economic theory will receive less attention. Economists will be valued less on their erudition and more on their creative capacity. The advances will be achieved not only by conceiving new hypotheses, but also by compiling statistics from fresh fields, by inventing new technical methods, by refining upon old measures, and perhaps by devising experiments upon certain types of behavior. It will be harder for anyone to cover the whole field, perhaps quite impossible.[35]

The inevitability of technical developments in methods and terminology as a science matures must be faced by social scientists as it has been faced by other scientists. It is no longer assumed in the mature sciences that every scientist, not to mention the layman, should be able to read with ease the technical monographs of every other scientist. The tradition is still strong in the social sciences, however, that if the layman or the literary fraternity finds a monograph dif-

[35] *Ibid.*, pp. 6, 7.

ficult or unreadable, that is fairly conclusive proof of its worthlessness. The journalist and the social philosopher with reputations gained during the preceding generation are sometimes among the first of those who deplore any new developments which they cannot follow without hard work.

The reaction of Hobbes to Wallis' work on Conic Sections (to take no more recent example) is illustrative. "Instead of describing geometrical figures and proofs in Latin sentences as had been the invariable custom previously, Wallis introduced the innovation of lettered diagrams for which students since his time may well be grateful; but Hobbes sent [the manuscript] back with the comment that he knew not whether the proofs were good or bad for it was so 'covered o'er with a scab of symbols' that he would not even read it!" [36] In saying that he knew not whether the proofs were good or bad, Hobbes took an entirely proper position and was remarkably frank and honest; the more common reaction would have been for him to have declared the proofs bad, because he could not read them, though this reason would usually not be admitted. Every social scientist may reasonably expect some day to be left behind in one or more of the fields of a rapidly advancing science. It requires moral courage and scientific humility to accept the position of a novice in a field where one is accustomed to being regarded as an authority. But if we do not wish to obstruct the advance of science, such relinquishment of authority and leadership at the proper time is required of the scientist. It is proper for us to maintain a healthy skepticism and a noncommittal attitude toward the developments which we do not understand. But we cannot safely condemn innovations merely because they are new, or because we find them difficult, or because we do not understand them.

We have discussed above some general characteristics of the scientific method as an approach to the solution of scientific questions. Before proceeding with a consideration of the more concrete techniques involved in this method as applied to social phenomena, it may be well to review briefly some considerations which influence the selection of questions to which the method is finally applied.

[36] S. C. Dodd, *op. cit.*, ch. 1.

L. THE SELECTION OF RESEARCH PROBLEMS

In the most general sense a research problem is determined by the circumstances which arouse in someone a curiosity which craves satisfaction. Curiosity in its most general sense is merely the "intellectual" (*i.e.*, symbolic or linguistic) aspect of a tension in the organism—some imbalance between it and its environment or between the parts of an organism. Thus scientific research is, in the last analysis, merely a more or less refined and abstract type of the same general activity which characterizes all life in its gropings toward adjustment within itself and with its environment. We reserve the name research, however, for only the more formal, deliberative and "intellectual" (*i.e.*, symbolic and abstract) forms of this general adjustment process. Scientific research, furthermore, is usually thought of as an activity of *general* import to the community, rather than as a struggle on the part of the individual scientist to achieve a personal satisfaction of his curiosities, although the latter may be foremost among the scientist's incentives.

The question whether the scientist should be primarily concerned with the solution of problems of more or less immediate concern to the contemporary world or whether he should select his problems primarily to satisfy his own intellectual curiosity is a matter of frequent controversy. Actually, of course, the two interests are likely to coincide to a large degree unless the scientist is too sheltered and removed from the life of his community and his generation. But it is frequently difficult to show the layman that some types of scientific research are relevant to contemporary community problems. In such cases the scientist frequently takes the position that he is a privileged person who must be allowed to follow "truth" or curiosity for "its own sake," unfettered by the current interests or needs of contemporary life. It is correctly pointed out that some of the most important discoveries in science have resulted when men followed merely their curiosity without thought of the possible practical consequences of their work. It is doubtless true, for example, that the mathematical researches of Faraday and Clerk Maxwell seemed to everyone at the time to be quite devoid of

practical application. It is also unquestionably true that these researches have been found subsequently to be entirely basic to our development of the whole field of radio physics. The same could be said regarding early experiments in electricity and many other fields. Finally, it will be recognized by everyone that, if research should become too preoccupied with the day-to-day problems of each community, the *generally valid knowledge,* which constitutes the great achievement and content of science, could probably never develop. There would be no time for such "impractical," long-run projects. There is real danger that an overly-practical and short-sighted policy in this respect might dry up the very fountainheads of scientific advance through a failure to support adequately the more "esoteric" researches into theoretical and mathematical questions which have always been the conditions of progress in science.

In times of public excitement especially there are likely to be outcries from "practical" men, reformers, and even some poets, as proponents of "action" against "lecturing on navigation while the ship is going down." It is because some men went about their proper business (instead of rushing at odd jobs about the ship, which could not be saved, or was not worth saving anyway), that we know something about navigation today. The researches and the lectures survived and were found useful on other ships on which, however different from the ships that sank, men found life quite tolerable and navigation greatly improved.

Past experience, may be said, then, to justify the maintenance by society of a class of people of certain characteristics and training, commissioned to do nothing but follow their "pure" or "idle" curiosity, on the ground that the apparently purely personal and frequently esoteric pursuits of these men occasionally yield, as an incidental by-product, results of vast significance to the community and to mankind. It is important to note, however, that when we make this concession we do so not because of any assumed "right" of the intellectual to such privileges, but purely on the *demonstrated probability* that this policy and practice will yield a concrete return, however indirectly, to the community. In short, we may resolve the controversy between the con-

tenders for "practical" research and the "truth-for-truth's-sake" contingent somewhat as follows: The "intellectual" is on sound ground when he stresses (*a*) the need of freedom in science to follow curiosities which have no known relation, or at least a very remote and hypothetical relation, to the problems that may be agitating the contemporary world; (*b*) the indirect significance to "practical" problems of such freedom and of highly "esoteric" mathematical and other research in the fields of symbolic representation; and (*c*) the justifiability of the long view in science, that is, the contemplation of, and work upon problems in the perspective of centuries rather than in terms of individual lifetimes. Scientists and other intellectuals are, on the other hand, on unsound and unrealistic ground if and when they contend for freedom to follow their curiosity, or "truth for truth's sake," *as an end in itself rather than as a means to ends in which the larger community has a stake.* Unless the intellectual can make out a reasonably probable claim that his mysterious activities will *someday* yield a return *of a kind that the community values,* he has no more claim to support than any other member intent only on his own amusement or development.

But science is perhaps in greater danger from overemphasis on "practical" research (in the sense of research on local and immediate problems) than on "pure" research. It is comparatively easy to get a hundred thousand dollars for a social survey of a local community if immediate results of value are thought probable. It is very difficult to get a fraction of that sum for research on the development of social measuring instruments, for example. Yet the latter, when developed, become the principal means of reliable and economical "practical" research as well as the chief means for the advance of science. The matter can perhaps be summarized most briefly by imagining what would be the state of knowledge, research, and practical efficiency in the physical sciences if every researcher had to invent anew for each research project, instruments for the measurement of time, length, weight, temperature, etc. The corresponding state of affairs is virtually what today exists in the social sciences. There is no systematic theory which stipulates what we choose

to regard as fundamental dimensions of social behavior and there are few standardized and verifiable definitions and measures of the various concepts and dimensions used by various researchers.[37] Deliberate and coordinated work on this key problem might properly engage some of the attention of research foundations interested in fundamental research.

M. PLANNING A RESEARCH PROJECT

Turning now from the above more or less theoretical aspects of the selection of research problems, we may consider some of the more practical considerations confronting the student in the selection and planning of a research project.

If a person is employed as a professional researcher, his employer will tell him to what questions answers are desired, on the basis of the needs of the industry, the government department, or the research institute in which the research is to be carried on. The same may be more or less true in academic research, when the director is interested in testing some broad theory through a series of graduate theses. If the problems so proposed are strategic or crucial to the further advancement of systematic knowledge in a field, the serious student will perhaps prefer to work on such a subject rather than try to find some unique topic which has hitherto escaped the interest of everyone else, perhaps because of its triviality.

The thoroughly absurd notion (perhaps adopted from historical research, in which field it may have some validity) that graduate theses should be on some new topic not yet touched by other researchers, has resulted in the scattering of the research energies of thousands of graduate students among a vast variety of unrelated topics bearing on no central scientific theory at all. It is true that the principal purpose of graduate theses is to provide research training for the student, and that this may be best achieved by allowing or compelling him to formulate his own problem. This is still

[37] This work will necessarily be, to begin with at least, heuristic in nature, *i.e.*, different students will propose different theories, concepts, and measures of which the most useful will survive in whole or part. An example of one of the most comprehensive and systematic of such attacks is found in S. C. Dodd, *Dimensions of Society* (Macmillan, 1941).

possible, however, within the vast province of conflicting theories and approaches which characterize the social sciences today. It is not a question of confining research to *one* theory. It is rather a problem of making graduate theses bear on *some* theory so that knowledge may grow by *systematic accretion.* The alternative is the mere accumulation of random fragments. For the same reason students should be encouraged to undertake theses on the *same* questions which others have already studied in order to corroborate or refute the previous findings. Only when the findings of one researcher have thus been repeatedly tested can we regard the results as reliable. The undeveloped state of clearly integrated theoretical frameworks in the social sciences is, of course, chiefly responsible for the chaotic scattering of research in these fields. As a science matures this uncoordinated effort tends to be supplanted by more systematic research directed at more clearly defined goals.[38]

The skilled researcher should be able to proceed effectively on any significant problem, whether he is motivated with great personal curiosity in the matter or not. As a technician he should find sufficient motivation in the purely technical problems of finding reliable and demonstrable answers to the questions asked, even if he does not consider the latter of great importance. The main problems of carrying on a research project are largely the same regardless of the precise character of the question to be answered, and we shall therefore consider briefly some of these problems.

1. When the general subject to be studied has been decided upon, whether through the researcher's own interest or through assignment by others, it is important to define and delimit as precisely as possible the aspect or aspects to be studied. This can best be done by contemplating exactly what kind of *answer,* in the sense of form and degree of specificity, is desired. This does not mean that the specific result itself can or should be decided in advance. Above all, it must be remembered that in research, negative data may be as important as positive data. Many research projects fail largely because of lack of clearly defined objectives. For

[38] The beginnings of a more systematic approach are becoming apparent in psychology. See C. L. Hull, *op. cit.*

example, after a year or two of research on many such topics as "child life," "women in industry," "leisure," "the newspaper," etc., it is found that, while a vast amount of "interesting material" has been gathered, practically no important scientific questions of general interest can be answered.

Every topic has an unlimited number of aspects and could be studied for unlimited periods of time. When a topic is suggested for research, each person usually has in mind *some aspect of it* to be studied. Others may have other aspects in mind. If each takes the broadly stated topic to *mean* the aspect in which he is interested, great confusion and disagreement in the appraisal of the results is likely to ensue. The researcher, especially the beginner, would do well not to undertake to investigate a wide, general topic but insist on an *explicit statement* of the *aspects* to be investigated. This can frequently best be achieved by a list of *specific questions* to which answers are desired. These questions become the guide to the planning of the investigation in all its details, including, for example, the types of tabulations to be employed. If a certain type of question is to be answered, a certain type of table must be forthcoming from the data in order to answer such a question. In order to produce a certain type of table, the schedule on which the data is recorded must be so constructed that the data necessary for the required table will be collected in the *necessary form*. This, in turn, as we shall see later, has everything to do with determining the categories and units in terms of which the data are to be gathered. Nothing is more wasteful and futile, or more characteristic of the inexperienced researcher than the tendency to engage in an enthusiastic gathering of data on the theory that, if only enough and sufficiently varied data are collected, any or all questions can be answered from the results. The researcher will find to his grief that too frequently his data can be used for the answer of even some of the simplest questions only if these questions were definitely in mind when the data were gathered. A question, then, serves as a framework for an investigation, and as every question implies an answer, it requires the researcher to envision at the outset the *form* of the answer which will satisfy the question with which all inquiry begins.

The more specific and definite that question is, the more efficient and specific the investigation is likely to be.

This does not mean that *broad and general objectives* may not be *specifically set* as the purpose of a research especially by mature students or to survey a field of problems. There are occasions where such preliminary or exploratory studies are desirable and necessary in order to formulate the more specific problems that everyone vaguely feels are involved. The degree to which such preliminary work is necessary will depend greatly upon the amount of research which has already been done in the field and upon the researcher's familiarity with this work. In general, it may be said that students as well as more experienced researchers are prone to undertake subjects unintentionally too broad in scope and out of all proportion to the researcher's resources in time and money. When the objectives are intentionally broad and general, it is important that even these objectives should be *deliberately,* planfully so, if the research is to proceed efficiently.

2. When the subject of a research has been decided upon and defined as precisely as possible, it is necessary to make some preliminary inquiry as to the feasibility of the undertaking from a number of points of view. First, the accessibility and availability of the sources must be considered. No matter how excellent may be the research decided upon, if it depends for its fulfillment upon sources which are not available within the resources of the researcher, the project is not feasible. To determine this question, the researcher must have decided at least roughly what factors the problem involves. Second, the relationship of the contemplated project to other projects, and especially its dependence upon such other projects, needs to be carefully considered. To the extent that the successful prosecution of the study is contingent upon factors outside the investigator's control, the undertaking must be considered precarious. Finally, there are questions of the practical purposes and results of the project: (*a*) To whom and in what respects will the results be of interest? (*b*) Will the results bear upon new hypotheses or will the findings mainly support or refute hypotheses for which there is already much evidence? (*c*) Are

the results intended for purely scientific purposes, or as data for various immediate and practical uses? Any or all of these may be significant and valuable contributions. But it is important to consider a proposed research from all of these points of view because the methods to be employed may depend to a large degree upon these considerations — for example, the degrees of accuracy and verifiability required, public reaction to the undertaking, etc.

3. Practically all research must go forward under fairly specific limitations of time and financial support. There is a more or less romantic view of research which paints the creative worker as one who, decade after decade, pursues the elusive problem he has set himself regardless of any other considerations. These conditions are approached in the cases of well established sciences and endowed research institutes and laboratories, and they are, of course, the ideal conditions for the solution of fundamental problems. But even in such circumstances, periodic reports and budgetary estimates are called for. The bulk of research, however, and notably social research, takes place under much less favorable conditions. Much of it is undertaken for administrative guidance and must be completed within definite time limits.

The conditions under which social research is carried on are so varied as to prevent here any detailed or itemized sample statement of budgets, either of time or money. The tendency for both to exceed their advance specifications should, however, be recognized and allowed for to the extent of at least five percent of the total budget. Accurate estimates of time and money required to carry on and complete a study depend largely on the possibility of itemizing carefully the time, equipment, and assistance which will be required. Among the items that usually must be included are: (1) Salaries of workers and expenses of meetings and conferences, if any. (2) Rental of office and equipment, desks, chairs, stenographic and statistical equipment, telephones, filing cabinets, etc. (3) Supplies — papers, stamps, printed supplies, schedules, books, maps, etc. (4) Traveling expenses. (5) Special costs, if any — *e.g.* mechanical tabulations by an outside agency, drafting of charts, etc. (6) Publication and distribution of the report.

The proportions of a budget to be allocated to each of these items will vary enormously with different types of projects and cannot be suggested in general terms. The researcher's own experience, and that of others who have conducted similar projects, should be consulted. The investigator's familiarity with the circumstances under which the study is to be conducted will be his most valuable aid in budgeting both time and money. In general, it may be said that the inexperienced researcher especially is likely to allocate too large a proportion of his time and money to the gathering of the data. Later he finds himself handicapped for lack of both in the analysis and presentation of the results.

4. The techniques of dealing with the problems that arise in the course of a study, especially in its first stage, namely, the planning of the project and the gathering of the data, are the principal theme of the remainder of this book. The more generally accepted methods of analyzing the data collected have been extensively dealt with in a large number of general treatises on statistics. The preparation of reports and manuscripts has likewise received adequate attention in a number of places to which references are made at the end of this chapter. Illustrations and further elaboration of many of the considerations here raised in outline form will occur throughout the volume.

N. SUMMARY AND CONCLUSION

In the foregoing pages we have taken the view that of the various methods which man has employed in his age-long struggle to adjust himself to his environment, there is one method, the general type we have called the scientific method, which has proved itself incomparably superior to all others in the results it has achieved. The implication seems clear that a method which has proved itself so effective in one field of human adjustment should be employed also in those fields where man's adjustment and control is relatively imperfect, namely, in his social relations. But up to the present, man has made only feeble attempts to apply scientific method to his social problems. It has been erroneously assumed that the data of human society are so

fundamentally different from the data of the physical world and of lower organic life as to make the method inapplicable to human relations.

This general objection, we find, reduces itself into a number of factors, all of which are more or less generally regarded as fatal obstacles to the formulation of true scientific laws of social behavior. Thus it is contended that the data are too complex. This is a subjective attitude toward things which we have not yet analyzed and hence do not understand. It is a condition which flows from the very inadequacy of our present description and definition of social phenomena.

Second, it is contended that social phenomena are subject to incalculable whims due to the alleged presence in human beings of "self-active" brains, "will," "minds," "souls" and other mystical concepts of an old and rapidly disintegrating psychology. We have attempted to show that these whims are present in "natural" phenomena as well,[39] and that scientific law meets the situation by standardizing the conditions or abstracting the situation so as to describe behavior under these conditions with a high degree of accuracy.

Finally, there is the fear that, since scientific method tends to be quantitative in its description, such description of human relations will omit various alleged intangible qualitative conditions which must be considered. A variation of this objection is the contention that a quantitative analysis of a situation destroys the *relationship* between the parts, the vital thing in society. We have attempted to show that there is reason to believe that the essential qualitative differences are all reducible to quantitative statement. As for the contention that quantitative analysis and statement do not take into consideration the *relationship* of the parts, this is true only of an imperfect or inadequate quantitative statement. A mathematical formula describes not only the constituent elements of a complex, but their relevant relationships. It is not contended that at present all the relationships betwen social phenomena are susceptible to statement in terms of mathematical or other symbols employed in the

[39] *Cf.* S. and B. Webb, *Methods of Social Study,* (Longmans, 1932): "We think that 'wert' and 'sinn' [value and meaning] enter no more and no less into one department of our scientific knowledge of the universe than into another." (p. 12)

physical sciences. To conquer new fields, new concepts and tools must be invented. They must be developed and improved through experience as they have been developed in the mature sciences. But we must seek these concepts and tools in the field of the quantitative and objective, rather than in the domain of the qualitative, subjective, and metaphysical.

In this evaluation of the objections to the possibility of exact social sciences, there is no desire to minimize the difficulties involved, nor to claim that much progress in their solution has as yet been achieved. But it is important to distinguish between great difficulty and impossibility. In view of the present admittedly low development of anything properly designated as science in the social studies, it may be objected that the position we have taken is unjustifiably optimistic. Our optimism is justifiable, however, in the light of the history of science. Every extension of the field of science has met with exactly the same objections as are today urged against the possibility of scientific method in sociology: The prospects of exact social sciences and a consequent transformation of the social world comparable to that which the physical sciences have wrought in the physical world, involve a far lesser imaginative leap than the scientific achievements of today would call for on the part of the scientists of several centuries ago.

Considerations of this kind are not efforts to make light of the difficulties confronting social scientists. These difficulties are very real; they will require years of patient and discouraging effort to solve. Some of the more obvious of these difficulties are considered in the following chapter. But while these difficulties will be emphasized, this should not be interpreted as a counsel of despair because of the impossibility of attaining in present social research some of the ideals of mature science that have been reviewed in this chapter. In fact, the student should be warned against an unreasonable quest for perfection that is unattainable. He who seeks complete knowledge and final accuracy in every detail will presently be eligible for the psychopathic ward, and, in any case, will never reach his goal or finish his work. This book will continue to emphasize the importance of accuracy, ade-

quacy, and caution in the gathering and interpretation of data. Yet even these watchwords are relative to the ends sought. For practical purposes, there should be in research, as in most other activities, a compromise between (*a*) what the researcher knows to be the ideal and (*b*) the practical requirements of the project. This compromise should be made at a point (*a*) where the results are sufficiently good for the purposes in mind, (*b*) where the results have *some* value even though imperfect, and, in any case, (*c*) where, under the circumstances, one has done the best one could. In all instances, however, it is important to *recognize and call attention to* the imperfection of the work which, for whatever reason, we are unable or unwilling to carry on to higher degrees of perfection.

For the same reasons we have recognized the value of a large body of literature which can hardly be said to conform to the formal requirements of scientific research but which is invaluable as sources of hypotheses and clues. These may suggest both the formulation and the interpretation of more rigorous scientific studies. The scientist has, therefore, no quarrel with "literature" from the most impressionistic poem, or novel, or life histories to the more objective descriptive accounts of ethnology and community studies of a purely literary sort. Neither does the scientist confuse such material with scientific findings. It is doubtless true that the ultimate aim of scientific research is to arrive at scientific laws. But it is also true that the bulk of social research aims not directly at this end, but rather at immediate objectives of administration and community adjustments. The problems and methods discussed throughout the remainder of this book will be found to apply with equal validity to both types of research.

O. SUGGESTIONS FOR FURTHER STUDY

1. The best training in scientific research is to participate in such work under the guidance of a capable scientist. The next best training is to be secured through a careful examination of the printed accounts such men have left regard-

ing the details of their labors. The reading of the present text should, therefore, preferably be accompanied by participation in an original research project, in the course of which the student may experience first hand the problems here discussed: arriving at a decision on the problem to be studied, formulating hypotheses, planning observations, and collecting data. (The analysis of the data belongs perhaps more directly in a course in statistics.) Whether or not such a project is undertaken, the student should carry on from the beginning of the course the study of one or more model research monographs. These will illustrate the way in which competent researchers have wrestled with similar problems in some specific cases. Such monographic studies will doubtless be recommended by each instructor according to his own judgment of the student's interests and maturity. The following are a few excellent examples.

(*a*) S. C. Dodd, *A Controlled Experiment on Rural Hygiene in Syria,* (Oxford Univ. Press, 1934). Pages 1-38 of this monograph have never been surpassed as a model of clear orderly statement of the methods of a research project. The few mathematical statements may be skipped by those who are not familiar with elementary statistics.

(*b*) E. W. Burgess and Leonard Cottrell, Jr., *Predicting Success or Failure in Marriage,* (Prentice-Hall, 1939). Chapters 1-5 are a very fine illustration of the problems and methods of approaching one kind of sociological problem. The method is applicable to a wide range of problems and should be carefully studied. Other studies of the same general type are (*c*) G. Vold, *Prediction Methods and Parole,* and (*d*) E. Monachesi, *Prediction Factors in Probation,* (both published by The Sociological Press, Hanover, N. H., 1931 and 1932, respectively). An excellent non-technical monograph for beginning students is (*e*) R. S. and H. M. Lynd, *Middletown in Transition,* (Harcourt, 1937).

2. Elaboration of the position taken in this chapter regarding the nature of science, the applicability of quantitative methods in the social sciences, and the nature of scientific laws, may be found in G. A. Lundberg, *Foundations of Sociology,* (Macmillan, 1939), chs. 1, 2, and 4, respectively.

3. Regarding the purposes and types of problems to which

social scientists should devote themselves, read R. S. Lynd, *Knowledge for What?* (Princeton Univ. Press, 1939), chs. 1, 4, and pp. 249-50. See also my review of this book in *Amer. Jour. Sociol.*, 45:270-74, 1939.

4. On the subject of analysis and interpretation of data the following texts, among many others, provide an elementary treatment, involving no advanced mathematical knowledge. The briefest, and, considering its brevity, the clearest summary of the whole subject is by Calvin F. Schmid, chs. 11 and 12 in Pauline V. Young, *Scientific Social Surveys and Research,* (Prentice-Hall, 1939). The student entirely unfamiliar with the field, however, will probably require a more elaborate treatment such as is to be found in the following: For students of economics and sociology, R. E. Chaddock, *Principles and Methods of Statistics,* (Houghton Mifflin, 1925); for students of education and psychology, H. E. Garrett, *Statistics in Psychology and Education,* 2nd Edition (Longmans, 1937). The most comprehensive general treatise is F. E. Croxton and D. J. Cowden, *Applied General Statistics,* (Prentice-Hall, 1940).

5. Suggestions on the preparation of reports and manuscripts may be found in W. E. Spahr and R. J. Swenson, *Methods and Status of Scientific Research,* (Harper, 1930), chs. 12 and 13. See also *A Manual of Style,* (University of Chicago Press, 1935); and W. L. McAtee, "On Scholarly Writing and Critical Reviewing," *Scientific Monthly,* July, 1940, pp. 77-79.

6. For a thorough discussion of the prediction of human adjustment, see Paul Horst et al., *The Prediction of Personal Adjustment,* Social Science Research Council Bulletin No. 43, 1941.

CHAPTER II

DIFFICULTIES OF OBJECTIVE OBSERVATION; EXPERIMENTAL TECHNIQUES

In the most advanced communities, and even among the adepts of modern science, there comes up persistently the revulsion of the native savage against the inhumanly dispassionate sweep of the scientific quest, as well as against the inhumanly ruthless fabric of technological processes that have come out of this search for matter-of-fact knowledge. . . Furtively or by an overt breach of consistency, men still seek comfort in marvelous articles of savage-born lore, which contradict the truths of that modern science whose dominion they dare not question but whose findings at the same time go beyond the breaking point of their jungle-fed spiritual sensibilities. — *Thorstein Veblen.*[1]

A. THE NATURE OF KNOWLEDGE

Our knowledge of the world must be secured through our response organs, commonly referred to as our senses.[2] The correctness of our knowledge, that is to say, the similarity of our perceptions when compared with other people's, depends in the last analysis upon the physical soundness of these response organs. But, unfortunately, and unlike a telescope or other mechanical tool of observation, the mere mechanical perfection of our response organs does not insure uniformity of perception among different observers. This differential in the perceptions of different persons when exposed to the same stimuli derives from several sources: (1) the perception of all but the simpler and cruder stimuli is the result of the training and conditioning of our response organs; (2) our responses are influenced by physiological and environmental conditions — fatigue, age, temperature, etc.; and (3) the reception and interpretation of the stimuli received is determined by our "apperceptive mass," that is, the sensitivity we have developed and the responses we have learned, as a result of past experience. "Men observe with the eye of the past." [3]

[1] *The Place of Science in Modern Civilization,* (Viking Press, 1932), pp. 26, 27.

[2] For elaboration of this statement, see *The Philosophy of John Dewey,* (P. A. Schilpp, Editor. Northwestern Univ., 1939), p. 538. See also J. F. Markey, *The Symbolic Process and its Integration in Children,* (Harcourt, 1928), pp. 14-15.

[3] Columbia Associates in Philosophy, *An Introduction to Reflective Thinking,* (Houghton Mifflin, 1923), p. 24.

B. THE NATURE OF BIAS AND PREJUDICE

The neuro-muscular set (symbolic mechanisms), which thus determines what shall be perceived and how it shall be interpreted is frequently referred to as bias or prejudice. In the case of the latter term, at least, the implication is that it is in some measure a weakness and a wicked thing, springing from motives of self-interest or other selfish considerations. This attitude toward a perfectly natural phenomenon represents, of course, the survival of an antiquated psychology and of theology. While the practice of imputing ethical and "intellectual" motives to the reactions of the rest of the animal world has been abandoned for some time, it still persists in regard to man. Thus, to use an illustration from Raymond Pearl: "An amœba behaves quite differently in going after his food according to whether this food is capable of running away, or is something like an algal cell that naturally stays quiet. In the former case the amœba makes provision in advance to prevent the prey's escape; in the latter it closes in at once without any preliminaries. When a man acts in this way, behavior is commonly called 'intelligent' or 'shrewd'." [4] It should be added that such behavior on the part of a man is also "wicked" if the object of his approach is socially taboo. To the behaviorist, bias and prejudice, being merely the more or less permanent conditioning of our nervous system with respect to certain stimuli, is, from the scientific standpoint, as devoid of moral significance as the behavior of the amœba. It is simply a normal property of protoplasm to have its future behavior tendencies conditioned and affected by its past behavior. Says Bernard: "All of us undoubtedly receive permanent organic and emotional and intellectual biases in our early years which we never fully overcome. Indeed it is not abnormal for us to do so." [5]

Reduced to its simplest form then, bias may be said to be a conditioned response which predisposes an individual to make his observations fit a previously established system of evaluations and habituations.

[4] Raymond Pearl, "The Biology of Health," *Amer. Mercury*, 6:386, 1925.
[5] L. L. Bernard, *Introduction to Social Psychology*, (Holt, 1926), p. 219.

C. THE PREVALENCE OF BIAS AND PREJUDICE IN THE SOCIAL SCIENCES

The fact that the perception and response tendencies called bias and prejudice in human beings are a natural and inevitable condition, does not reduce its importance as a problem to be reckoned with in the employment of the senses in scientific observation. Bias and prejudice are complicating factors in all sciences, but their importance is much less in the physical than in the social sciences. The chief reason for this is that the subject matter of the physical sciences tends to be more removed from the common emotional complexes which we develop toward many social phenomena; and hence the perceptions by normal senses of physical facts tend to be fairly uniform. As Keller has said: "A man can count the legs of a fly and report his findings without having his heart wrung because there are too many or too few."[6] But it is very difficult for the person who dearly loves his cocktail to induce himself to observe and report that prohibition seems to have been an economic boon to the community which he is studying. Likewise there is the example of the eminent historian who, on being confronted with various documentary facts tending to throw much doubt on his preconceived opinion as to responsibility for World War I, made this startling admission:

> The subject is too involved, the underlying race and language antipathies are too strong, the confusion of relations in Eastern Europe too complex *to make any review of printed testimony a safe basis for changing an opinion which was forged by the fires of war.*[7] [Italics ours.]

D. INADEQUACY OF SUBJECTIVE JUDGMENTS

While this tendency to observe and interpret data according to our prejudices is especially great in considering social phenomena — matters on which we have strong emotional attitudes of love, hatred, reverence, etc. — it is by no means

[6] "Sociology and Science," *The Nation*, 102:475, 1916.

[7] Albert Bushnell Hart, *Current History*, 20:196, 1924. Comment on an article by Harry Elmer Barnes on "Assessing the Blame for the World War," *ibid.*

confined to the field of social relations. We have a certain acquired sense of the "fitness of things," "reasonableness," and "common sense" which frequently causes us to refuse to accept a conclusively demonstrated proposition even when the data involved are of purely academic interest and do not touch our lives in any way. For example, there is this familiar hoax:

> Suppose an iron ring was placed around the earth at the equator, assuming the surface to be smooth so that a close fit could be obtained. Suppose the ring were cut in one place and seven additional feet were inserted into the circumference. How far would this cause the hoop to stand out from the earth?

When people are asked to give an impressionistic guess as to what effect this increase in circumference would have on distending the circle, the practically universal response is that the difference would be "imperceptible," "enough to put a very thin paper under," "one thousandth part of an inch," etc. And yet it can be demonstrated by a process known to every schoolboy (the relationship between circumference and diameter of a circle) that the hoop in question would have its diameter increased by approximately two feet as a result of the addition of seven feet to its circumference, thus causing the hoop to stand out at least a foot from the surface of the earth all around. But this easily demonstrated conclusion is usually rejected as preposterous on the spur of the moment even by scientists and mathematicians. The only reason they can assign for refusing to accept the inescapable conclusion is that it "doesn't seem reasonable." All that this means, of course, is that the results do not seem to conform with their sense experience within a realm very much more circumscribed and subject to direct and unaided sense impression.

E. THE "COMMON-SENSE" FALLACY

Psychological and statistical treatises are full of examples of the enormous and ludicrous errors of our senses, unaided and uncorrected by objectifying devices outside of our organic equipment. Nevertheless, the great majority of people, including most of the so-called educated, will without hesitation prefer to follow what they call their "reason" or

"common sense" when the conclusion reached through these much-revered faculties clashes with the findings of instrumentally corrected sense impressions and logic. Herein lies one of the most important tests of the true scientist: has he the power coldly to set aside his preconceived subjective notions of the outcome of an experiment or investigation when the carefully checked objective data lead to other conclusions? It is perfectly desirable and defensible for him to use his subjective impressions and his empirical conclusions as a basis for skepticism of the conclusions reached by objective methods, especially when such skepticism leads him to repeat the experiment or to recheck his data. For this purpose, "common-sense" and "reasonableness" will always be the criterion by which results will be judged. The mischief occurs when these subjective judgments are accepted as infallible, as substitutes, or as more reliable than objectively verifiable conclusions. The objective checks are, of course, not devices to be regarded as means in themselves of gaining knowledge of the facts. They are merely tools to assist and refine the senses when these organs are called upon to make observations to which they are unadapted, unaccustomed, or under emotional or other handicaps. As a matter of fact, the validity of a person's "common-sense" itself depends on how objectively that sense has been developed. It is only by virtue of objective external checks that our sense impressions have developed a validity even in the simpler adjustments of life, governed by "common-sense." The misjudgments of distance, size, and shape of a young child become corrected only by means of objective contact with the phenomena observed.

F. THE DEVELOPMENT OF OBJECTIFYING DEVICES

The recognition of this fallibility and inadequacy of the unaided senses has led to the invention of an impressive array of artificial devices for the objectification, standardization, and extension of crude sense impressions in the physical sciences. In fact, the development of these sciences has been very directly contingent upon developments in the tools whereby observations may be standardized. The requirement that scientific facts must appear the same to all normal

minds has made necessary (1) the development of objective devices whereby the senses of different persons can be made to perceive alike, and (2) the development of a standardized system of symbols and terminology by which sense experiences could be accurately described and communicated unambiguously to others. The second of these developments will be treated in the following chapter. Let us consider here the nature of the devices which the physical sciences have invented to standardize observation.

G. TYPES OF OBJECTIFYING DEVICES

The paraphernalia invented for the purpose of improving our observation of phenomena, and thus rendering more adequate the first step in scientific method are of three general kinds: (1) The devices invented for the purpose of securing a more definite delimitation and measurement of units of observation. In this category fall the scales, calipers, rulers, graduates, etc., without which physical science, and the arts resting upon them, would be capable of only the crudest generalizations and manipulations. (2) The devices invented for the purpose of projecting the senses or increasing their power. In this category fall such instruments as the telescope, microscope, and the microphone. (3) Devices which serve both of the above purposes in that they both intensify the powers of the senses and measure the phenomena in standardized units. Such instruments are exemplified by the thermometer and the barometer. Upon the invention and development of these devices for the projection of the senses and the standardization and objectification of sense impressions the advance of physical science has been directly dependent.

The importance of these external objectifying devices in the development of physical science suggests that corresponding devices must also be invented to correct and standardize the observations of social phenomena, if the same degree of accuracy is to be achieved in the social sciences. In fact, such devices are of even greater importance in the social sciences because of the greater natural tendency, already mentioned, toward bias and prejudice.

What progress, then, has been made toward objectifying and standardizing the observation of social phenomena and what are the possibilities in this direction?

H. OBJECTIVE TRENDS IN THE SOCIAL SCIENCES

While it must be admitted at the outset that there does not exist in social science any device of social observation or measurement comparable in nature or accuracy to the microscope or the calipers, the transition toward objective standards is definite. In the development of social theory, for example, Bernard sees "a more or less constant movement toward an objective statement of social problems. Hobbes' philosophy was a protest against the subjective and noumenal character of Scholasticism. Locke pointed out, in his own terminology, the relativity of subjective presentations and criteria. Bentham's professed and real purpose was to obtain a constant and scientific experimental basis for regulating morals and legislation, under which regulation everyone should be equal in privileges. . . His followers, John Stuart Mill, Spencer, the social or neo-utilitarian ethicists, and the early sociologists have constantly extended the analysis and have modified the criterion, till at last its hedonistic character is largely destroyed or disguised, though its subjectivism remains."[8] Likewise the transition from a purely historical to a comparative method in history and anthropology is a trend toward objectivism.

The tendency in recent times has been toward the inductive approach and the quantitative statement of social observations. This trend has brought with it as an essential counterpart, some devices, as yet crude, for the objective observation, measurement, and weighing (scoring) of social data. Up to the present these devices have taken the form of schedules, score cards, scales, and various other statistical devices. The greater part of the remainder of the book is devoted to the consideration of these all-important devices, upon the development of which largely depends the future of social science.

[8] L. L. Bernard, *The Transition to an Objective Standard of Social Control*, (University of Chicago Press, 1911), pp. 88-89.

I. INTELLECTUAL HONESTY AND THE SCIENTIFIC METHOD

Important as is the development of instruments and techniques of observation and measurement in science, it should not be assumed that either techniques or instruments are substitutes for an alert mind and common intellectual honesty. It sometimes happens that a scientist's greed for applause becomes greater than his devotion to truth. These are the strange cases in which a scientist will deliberately "doctor" his data in order to cause them to conform to a preconceived conclusion. This is plain cheating, and, when it occurs, it undermines science. Errors of this kind cannot be guarded against or corrected by instruments. The safeguard may be called an ethical one, and like all ethics, it is relative to the end sought. That is, scientific training includes training in certain ideals (standards of accuracy, veracity, etc.) as necessary means to the end sought. From this point of view, there is no need of introducing "ethics" as a separate category, since it is merely an aspect of scientific competence, without which scientific ends are imperfectly achieved. The competent scientist will therefore avoid deliberate corruption of his methods for the same reason that he avoids carelessness at any stage of his work. Since a corruption of the data is incompatible with the end sought, training in scientific method and in scientific ideals (standards) is perhaps the only safeguard against errors of this kind.

In those fields of observation in which objective measurement is not yet developed, familiarity with the psychological bases of prejudice is perhaps most helpful in enabling us to escape its warping influence. The person who is well acquainted with the etiology of his desires and emotions, the mechanics of their behavior, and their influence on his observations, will take both objective and subjective precautions to correct for their influence. Thus Darwin urged that it is well to write down facts contradictory to a particular theory that we wish to test, for otherwise we are likely to forget them.[9] Judges will sometimes decline to sit on cases with which their connection has been such as to make a bias

[9] W. F. Ogburn, "Bias, Psychoanalysis, and the Subjective in Relation to the Social Sciences," *Publ. Amer. Sociol. Soc.*, 17:66, 1922.

probable. Sometimes a person's consciousness of a bias is so strong as to cause him to "lean over backwards" in trying to correct it. While scientific measurement is the only adequate method of correcting the influence of bias, a knowledge of the roles which desire, rationalization, fantastic thinking, etc., play in influencing our observations will help us to avoid the effects of some of our cruder prejudices.[10]

J. THE NON-ETHICAL NATURE OF SCIENCE

A more common and less conscious source of error is the tendency of scientists, and especially of social scientists, to permit the current code of morals of the community or their own personal notions of ethics to influence them in their collection and manipulation of data. The findings of science are *per se* non-ethical. It is not the business of a chemist who invents a high explosive to be influenced in his task *as a scientist* by considerations as to whether his product will be used to blow up cathedrals or to build tunnels through the mountains. Nor is it the business of the social scientist, in arriving at laws of group behavior, to permit himself to be influenced by considerations of how his conclusions coincide with existing notions, or what the effects will be of his findings on the social order. For his purpose *as a scientist,* science must be an end in itself. *As a human being,* the scientist may properly look to the social results of his conclusions. While he may not be in any better position than any other intelligent and informed person to say what society should aim at, he should be invaluable in providing the most efficient technique whereby whatever ends are sought may be reached. But as a scientist his only ethical responsibility lies in seeing that the rules of scientific procedure have been complied with. Here again the only safeguard is thorough training in scientific method.

This does not mean that the social scientist may not as a scientist devote himself to questions of social welfare, however defined, or to other questions involving an appraisal or an evaluation with reference to an end. It merely means

[10] For illustrations of the operation of such factors see L. L. Bernard, *Introduction to Social Psychology,* (Holt, 1926), ch. 13. See also John Dollard, *Caste and Class in a Southern Town* (Yale Univ. Press, 1937), ch. 3.

that he should be clear in his own mind, and make it clear to others, when he is dealing with "what is" and with "what ought to be." Provided this is clear, the social scientist is not only justified, but is perhaps the only person qualified to specify what changes in existing conditions will bring about an elimination of conflicts or conditions which for any reason are regarded as obstructive to whatever end the community may (rightly or wrongly in the scientist's opinion as a human being) regard as desirable. From this point of view, the scientist who is dealing with "what ought to be" is only engaged in the scientifically legitimate practice of prediction, based on observations of "what is." He is merely predicting that a certain manipulation of factors in "what is" will result in something different. If he proceeds further and advocates the new product rather than the existing one, however, he is proceeding according to an ethical and not according to a scientific canon. There is no logic involved in science which requires him to take the last step. If he takes this step, he does so not as a scientist, but as a human being moved by an ethical motive.

There are those who feel that this interpretation of science reduces it to a cold technique devoid of human motivation. To them it might be pointed out that the pure scientist finds abundant motivation in the faith that only by a knowledge of "what is" can we make any practical programs of "what ought to be." Without this foundation our utopias are futile because they may be impossible or undesirable. In short, the scientist has ample motive in his belief that only by a more perfect knowledge of the nature of his universe can man adjust himself intelligently to it.[11]

K. THE EXPERIMENTAL METHOD IN THE SOCIAL SCIENCES[12]

The development of objectifying devices, as well as their accurate employment on the phenomena to be studied in

[11] For a more thorough discussion of this subject and the problem of "values" in the social sciences, see G. A. Lundberg, "The Future of the Social Sciences," *Sci. Monthly,* (Oct. 1941, 346-359). Also *Foundations of Sociology,* pp. 29-31, 527-28, 533-34.

[12] The material in this and the following section is largely an adaptation of Professor F. Stuart Chapin's pioneer contribution to this subject in his "The Experimental Method in Sociology," *Sci. Monthly,* 4:238-47, 1917. The quotations and most of the illustrations and analysis are drawn from this source, by permission.

the physical sciences, has been accomplished for the most part in the laboratory by the experimental method. Among the great advantages of this method are, first, that it permits the repetition of observations under practically identical conditions. This facilitates the verification of the observations by many observers. Second, it enables the observer to vary only one condition at a time, and to maintain all other conditions constant. This allows us to analyze the relations of cause and effect with much greater speed and certainty than is possible under uncontrolled conditions. For, if two conditions are varied at one time, it is not possible to tell which condition produced whatever effect occurred, or whether both acted jointly, or whether they neutralized each other. But under experimental conditions this difficulty is obviated by artificially controlling the variables. Because of these tremendous advantages of the experimental method, the question of its applicability in the social sciences becomes one of major importance. If this method is not applicable to social phenomena, it seems reasonable that the social sciences will be greatly handicapped, if not entirely thwarted, in their maturation.

The social scientist will probably never be able to bring into his laboratory a piece of society, there to observe its behavior in a test tube under varied conditions. But neither have astronomers ever brought the solar system into any laboratory. Their science has prospered, nevertheless. It is true that astronomical observatories contain very ingenious symbolic and mechanical representations of the astronomical aspects of the solar system and remarkable instruments for observing it. These, every science, and especially the social sciences, must develop. The experimental and laboratory techniques must in any case be adapted to the subject matter in which we are interested, and need not have much in common with those of physics and chemistry.

The particular technique of the experimental method varies widely in different fields. In biology, for example, we find it consists of the exposure of animals, as individuals and as groups, to varied environmental conditions and of the observation of their behavior. In psychology, the experimental method has been developed and applied to a great extent to

the behavior of human individuals, including their social behavior. The difficulties involved in such experimentation on large social groups are apparently formidable. But since the technique of experimentation must vary with the nature of the subject matter, and since very little attention has yet been given to the development of such a technique in the social sciences, the assumption that this method is not applicable to social phenomena is hardly warranted.

In fact, the use of the experimental method in various forms in the social studies is by no means unknown or even new. It has been extensively employed in the past and is being increasingly employed at present. Thus, the social studies have always availed themselves of what Chapin has called "natural experimentation." This consists of observing social behavior where nature has limited or held constant certain physical factors. The social facts observed in such areas are compared with those observed in other regions where these physical factors are or are not present. Of course, if we define experimentation as taking place only when there is actual *human* interference with the determining conditions, this natural experimentation is not experimentation at all. In any case, it is a very crude method because of the presence of many other variables than the general geographic and physical differences. But even with these limitations, the method has been of tremendous value to anthropologists and sociologists. For example, the isolation from the rest of mankind, the lack of variety of flora and fauna, and the climate have been taken to explain many of the customs and modes of living of the Eskimo, such as patricide (owing to the scarcity of food), the elementary organization of property rights, and absence of commercialism. Likewise, the survival of colonial customs in the Appalachian highlands of the south, while the rest of the country has been revolutionized, constitutes a natural experiment in the effects of isolation. While such studies have great suggestive value, they are not true experiments in that *the conditions are not subject to the manipulation of the observer.* As a result, there are too many varied factors present. Furthermore, as the number of similar cases available for observation are few, and since they are difficult to observe properly,

this must be regarded as only the crudest approximation to an experimental study.

L. THE UTOPIAN COMMUNITY "EXPERIMENTS"

Perhaps the best examples of avowed social experimentation are to be found in the Utopian Community experiments conducted by the associationists of the last century. In the establishment of such communities as New Harmony (1825) and Brook Farm (1841–47) the attempt was deliberately made to eliminate certain competitive conditions obtaining in society at large in order to achieve personal freedom, congeniality among people of minority opinions and even economic prosperity. Robert Owen, believing that the evils of the capitalistic system resulted from private property, orthodox religion, and the institution of marriage, established the community of New Harmony for the avowed purpose of testing his hypothesis by omitting these three institutions in their usual meaning. The community soon failed and broke up. On the assumption that these three fundamental human institutions were actually eliminated or reduced to constant elements, we have experimental proof of the instability of society without them. But unfortunately, the superficiality of the experiment allowed the presence of a number of other variables, each of which might, and, in fact, probably does, explain the break-up of the community. Thus, there was the heterogeneity of the population and a resulting "absence of like-mindedness," owing to the imperfect control of the experiment itself. Furthermore, there was a lack of isolation from the disturbing influences of outside society, as well as the varying social heritage and different value systems which the members themselves brought with them. People born and bred under other social conditions could not, of course, at once develop a new set of habits of thought and conduct. This inability to eliminate the prejudices, customs, and standards of outside society at once destroys the validity of the experiment. As Chapin has said, "It was like attempting to conduct a chemical experiment in a bowl of molasses."

The somewhat similar experiments of Fourier, while more successful than the New Harmony attempt, also failed to establish themselves on a permanent basis. But here again,

the social standards of outside society, to which members of the experimental communities were still susceptible, make it impossible to draw any conclusions as to the social value of the innovations tried in the communities themselves. The conclusion that it was the lack of isolation from the surrounding society rather than the innovations *per se* of the experimental communities which destroyed them is strengthened by the fact that where greater isolation has been secured, the experiments have sometimes succeeded. Thus, the Mormons of Utah inaugurated a social order in many respects unique, and flourished under it for many years. It is true that a new religious revelation was the essential condition in the success of this experiment, as in practically all others that may be regarded as successful. The religious factor is important in that it provides an accepted authority, thus minimizing dissension.

These experiments suggest that perhaps valid social experimentation is difficult because experiments must always take place in a social medium, the complicating effects of which we do not yet, at least, know how to calculate and allow for. The only way to avoid the complicating effects of the larger society surrounding a small community where an experiment is being tried, would appear to be to attempt the experiment on the larger society itself, that is, on the state or the nation. This, again, is by no means an unknown form of experimentation, as we shall see below. Sovereign states, through legislation, have long been engaged in experiments of this type.

M. SOCIAL LEGISLATION AS SOCIAL EXPERIMENTATION

Although legislation is not usually enacted for the purpose or in the spirit of experimentation, it is undoubtedly the most important type of social experimentation on a large scale in existence today. Legislation or enactments such as the Prohibition Amendment, the Woman's Suffrage Amendment, the government operation of railways during the first World War, the ventures into government ownership and insurance and much social legislation enacted under the "New Deal" 1933–36, are clearly modifications and innovations of an experimental nature. Overshadowing in scope and importance all such experiments, ancient and modern,

is perhaps the Soviet regime in Russia. The results of all such experiments, if accurately measured and reported, would clearly be of the greatest importance to social science.

The difficulty with this form of "experimentation" is that it is almost entirely of the trial-and-error type. Much important legislation is enacted every year and important social changes take place every year. But whether the legislation is causally related to the changes in the department directly aimed at by the legislation we have no definite method of determining. When the direct effects of the legislation are reasonably clear, these effects may themselves be the causes of other problems in other departments of society. Thus, a stringent divorce law may effect its immediate purpose, namely to cut down the number of divorces. But it may also cause an increase in illicit relations and illegitimacy. The trial-and-error process of legislation with individual unrelated statutes instead of scientific programs of legislation anticipating the effects on all other fields, is, like all trial-and-error experimentation, a blundering and wasteful process. But it is perhaps the only method possible in a new field, where the knowledge of the significant factors and their interrelationship is very limited.

Here we are confronted again with the principal difficulty of experimentation in the social sciences—the difficulty of controlling or calculating all the variables. This is due to two chief causes: (1) the complexity of the problem and (2) the undeveloped state of our methods of social measurement. The first cause has already been shown to flow from our ignorance of the factors involved. The difference between the trial-and-error method prevalent in social experimentation today (prevalent in all sciences in their early stages), and the highly controlled experiments of the physical sciences is that in the latter case *all the important conditions are known.* Furthermore, very exact methods of measuring these conditions have been developed. Hence their influences on the gross results of the experiment can be allowed for. The result is that the outcome of such experiments can be attributed with comparative certainty to a single variable. Frequently this result is attained by using "control" individuals or groups. Thus if a biologist desires to determine the effect

of a certain food on the physiology of rats, he selects two groups of rats as much alike as possible and maintains them as nearly as possible under the same conditions. One group is fed the particular food the effects of which the biologist is studying. From the other group this food is withheld. Careful measurement of the physiological conditions of the rats of the two groups reveal what differences, if any, are to be observed in their size, weight, or other conditions. The differences are then properly assumed to result from the factor present in the one case, but absent in the other.

N. SCIENTIFIC SOCIAL EXPERIMENTATION

The importance of accurate methods of measurement in social experimentation is further emphasized by the fact that perhaps our most successful social experiments have been in the field of public health. The fact that such objective measures as death rates are available in this field, and the fact that much is known regarding the conditions causing some diseases, enables us to measure the results of health programs with considerable accuracy. An excellent example of this technique is found in the Framingham Community Health and Tuberculosis Demonstration. The Metropolitan Life Insurance Company in 1916 placed $100,000 at the disposal of the National Association for the Study and Prevention of Tuberculosis to conduct an intensive experiment to determine whether it is possible substantially to reduce the morbidity and mortality of tuberculosis. First, a thorough study of existing conditions both of the people and their environment was undertaken. This included a study of the accuracy of the reporting of vital statistics, a study of general sanitary conditions, milk production and handling, a tuberculosis survey of cattle, food sanitation in stores, school hygiene, industrial hygiene, etc. The people themselves were studied with reference to the prevalence of physical defects, social and economic status and other objective measures of the existing conditions. All this material was carefully recorded in such form as to make possible rigid comparisons after the health program had been put in force. A comprehensive health campaign was then inaugurated. At the end of the first five-year period of the experiment the city showed a re-

duction of sixty-seven percent[13] of the tuberculosis mortality rate. The experiment is notable for the care with which the records were kept throughout the period and the allowance for possible variations in other conditions at the beginning and at the end of the period.

Foremost as an example of meticulously careful experimentation on entire communities is S. C. Dodd's study of the effect of education in hygienic practices in rural villages in Syria.[14] Especially worthy of note in this case is the care with which measuring instruments were perfected and the sample villages chosen, before the experiment was undertaken. Dodd divided the villages into two samples for experimental purposes and carefully measured, with specially prepared and tested instruments, the hygienic status and practices of each. One sample was then subjected to a two-year period of hygienic education through an itinerant clinic. At the end of this period the hygienic status of the two samples were again carefully measured. The comparison of the scores of the experimental and the control villages yielded a reliable measure of the results of the program of health education. It was found that (*a*) all of the villages improved significantly in their hygienic practices and conditions during the period considered, and that (*b*) the experimental village improved more, though the amount of the improvement was not statistically significant, probably because of (*c*) the difficulty of isolating entirely the control villages from the influence of the program of health education conducted in the experimental village. Nevertheless, the study is notable as a model of scientific proficiency in the care with which the measuring instruments were prepared, tested, and applied, as well as in the attention to problems of sampling and probable errors.

As an example of more restricted laboratory experimenta-

[13] National Association for the Study and Prevention of Tuberculosis, "Framingham Community Health and Tuberculosis Demonstration," *Framingham Monograph,* No. 9, 1922, p. 4.

For other notable examples of social experimentation, see Edgar Sydenstricker, "Statistical Evaluation of the Results of Social Experiments," *Jour. Amer. Statis. Asso.,* 23: (N.S.), 1928. Supplement, (Proceedings of the Eighty-ninth Annual Meeting), pp. 155-65.

[14] See the reference to this model study in the "Suggestions for Further Study" at the end of Chapter I.

tion on carefully controlled social groups, there is Lippitt's study of the effect of democratic and authoritarian group atmosphere upon the behavior of two children's clubs.[15] The clubs were formed after careful consideration of their comparability with regard to individual characteristics of their numbers, cultural homogeneity, position and relationship of members to each other, and to the larger social field, such as the class in school. In one of these groups the leader encouraged a maximum amount of choice and decision after discussion by the members on all matters. In the other club, the leader made these decisions for the most part and, in general, ordered the pupils about in authoritarian fashion. Extensive stenographic, descriptive "total behavior," and statistical records of the ensuing behavior were kept.[16] These results are rich in hypotheses as to the effect of different kinds of social "climate" upon social behavior and interpersonal relations. The study is notable for its attention to detail in setting up the experiment and for the variety and ingenuity of its methods of analysis. The sociometric and topological methods used in the latter connection are of special interest. These will be discussed in a later chapter in connection with the work of Moreno and others in this field. Moreno's numerous experiments on the social effects of changing the personnel of groups in the Hudson State Training School, as well as his experiments in psychodramatic work, are valuable as examples of the observation of social behavior under controlled conditions.[17]

The most careful and comprehensive experimental studies of the effect of varying conditions upon group behavior is to be found perhaps in the literature of industrial psychology.

[15] R. Lippitt, "An Experimental Study of the Effect of Democratic and Authoritarian Group Atmosphere," *Studies in Topological and Vector Psychology, I* (Univ. of Iowa, 1940). See also, R. Lippitt and R. K. White, "The Effect of Laissez-faire, Democratic, and Autocratic Social Climates." (Unpublished study.)

[16] In subsequent experiments moving picture records were also used. See R. Lippitt and R. K. White, *op. cit.*, (unpublished study, Univ. of Iowa). Reported in K. Lewin, *et al., Studies in Topological and Vector Psychology, I* (Univ. of Iowa, 1940), p. 107.

[17] J. L. Moreno, *Who Shall Survive? A New Approach to the Problem of Human Interrelations* (Nervous and Mental Disease Publishing Co., Washington, D.C., 1934). See also the journal *Sociometry*, which is largely devoted to experimental work in the field of interpersonal relations.

Mayo,[18] Whitehead,[19] and Roethlisberger and Dickson,[20] for example, report a large number of studies of experimental groups under carefully controlled conditions in a large industrial plant. While these studies are primarily concerned with the productivity of the worker, they throw a flood of light on the conditions governing, and the results of, personal relationships, incentives, freedom, self-direction, and morale.[21] Important as an example of refined laboratory experimentation also is the work of Chapple and Arensberg in the measurement of social interaction of human beings.[22]

O. OTHER EXPERIMENTATION AND REFINED OBSERVATION

As a rule the term "experiment" is, as we have seen, used loosely in the social sciences to indicate any innovation or carefully planned observation. The larger of these undertakings, such as the so-called experimental schools and colleges, have erred as scientific experiments by failing to keep sufficiently objective records of the conditions and the results of their educational innovations. That is, it has been for the most part impossible to attribute any variation in their results directly to the experimental conditions because of failure to control adequately or allow for other factors. More successful in these respects because of the more detailed observation and recording of actual behavior have been the laboratory observation of nursery school children with respect to personality traits and elementary social behavior. The following examples from the Institute of Child Welfare Research of Teachers College are illustrative of the type of controlled observation now being applied to human groups:

Floor plans were drawn of the roof and indoor playrooms where the nursery school children spend the first two hours of the day

[18] Elton Mayo, *The Human Problems of an Industrial Civilization* (Macmillan, 1933).

[19] T. N. Whitehead, *Leadership in a Free Society* (Harvard Univ. Press, 1936).

[20] F. J. Roethlisberger and W. M. Dickson, *Management and the Worker* (Harvard Univ. Press, 1939).

[21] For experimental attempts to measure morale, see L. D. Zeleny, "Sociometry of Morale," *Amer. Sociol. Rev.*, 4:799-808, 1939.

[22] E. D. Chapple with the collaboration of C. M. Arensberg, "Measuring Human Relations: An Introduction to the Study of Interaction of Individuals," *Genetic Psych. Monographs*, 22:3-147, 1940.

in relatively "free" (in the sense of minimally directed) play activities. In all these rooms, there is a variety of objects, some of which are large and stationary, *e.g.*, jungle-gym, swings, sandpile, piano, platform and steps, etc.; others may change from day to day, but remain stationary during any particular day, *e.g.*, rabbit cage, box on which nails are hammered, slide, etc.; others are carried about by the children at will, *e.g.*, blocks, doll-carriages, kiddie-kars, and innumerable small toys. The larger, immovable objects are indicated on the floor plans. Records each week have been obtained of the activities of each child of the lower age group by tracing his progress on these floor plans over a five-minute period. Every stop that he made was indicated, and timed, in seconds, with a stop-watch. His approach to and contact with other persons, adults or children, was noted by appropriate symbols, as likewise the contact or approach of other people towards him. Spontaneous activity was indicated, as was also "directed" activity. The time any particular object was in his possession was shown by indicating when he picked it up, and when he dropped it. . .

Another study on social behavior being made is closely related to the study by Miss Parten at Minneapolis. The technique here calls for the recording of every social grouping occurring during a given period, *i.e.*, every time two or more children are together. The time each child comes to and leaves the group is indicated. If possible, the "function" or part played by each child in the group is also indicated briefly, and a check is placed by the child who has initiated the group, if this is obvious. Probably the only data available for statistical purposes are those indicating the time spent by each child in a group, which will make possible the computation of an index of group participation for each child.

On the theory that laughter has a large element of the social in it, a study has also been made of the laughter situations in the nursery school. On a specially prepared blank, every laughter situation is recorded, with the names of the children in the group where the laughter occurred, and checks indicating which of the children "exposed" responded by laughing or smiling. Because of the interest inherent in this study of laughter, it is being made on three separate age groups. Thus, in addition to a laughter index (percentage of laughter to exposure) for each of the nursery school children, which will be used in comparison with the other social data, there is a genetic study of laughter in young children. Several thousand situations have been recorded, and these have been analyzed in their relation to chronological age, mental age, and other variables.[23]

[23] W. I. and D. S. Thomas, *The Child in America*, (Knopf, 1928), pp. 522-24. For more complete accounts of this kind of laboratory observation, see D. S. Thomas, A. M. Loomis, and R. Arrington, *Observational Studies of Social Behavior, I*, (Institute of Human Relations, Yale University, 1933).

These methods are largely dependent for their success on the development of tools and techniques of objective observation, recording, and measurement. The motion-picture camera with sound recording has vast possibilities, as yet largely unexploited, in this field. By means of these devices it is now possible to observe and record intricate social situations for subsequent repetition and detailed scientific study in the laboratory.

P. SOCIAL EXPERIMENTATION AND THE STATISTICAL METHOD

Outside of the laboratory, the method of controls is rarely applicable. In social experiments especially, it is usually necessary to experiment on a single group, and to draw our conclusions from a comparison of the significant facts *before* and *after* a new factor is introduced. This gives rise to serious complications owing to the lack of control of important conditions during the whole period of the experiment. But if we cannot *inject* or *eliminate* external conditions which may influence the outcome of an experiment, may not the same result be attained by *measuring the effect* of these influences? In this way we can allow for the influence of each variable in the gross results, and conclude that any residuary difference in social conditions before and after a new factor was introduced must be due to this new factor. If we keep a careful record of *all* significant social conditions, for example, before a law is enacted, and keep *strictly comparable* records after its enactment, may we not draw conclusions of some validity regarding the results of the experiment? If we cannot take into consideration and make allowance for *all* the conditions that influence an event over a period of time, which is probably impossible, we may be able to get a "first approximation" to the truth.

One of the commonest forms of such "experimentation" in the academic field is to give tests before a course is taken and to give similar tests at the end of the course.[24] It is assumed, although no rigorous attempt is made to allow for other possible influences, that the change in information and attitude, if any, results from the course to which the stu-

[24] See, for example, S. Menefee, "Teaching Sociology and Student Attitudes," *Sociol. & Soc. Research*, 22:545-56, 1938.

dent has been exposed. Menefee has also used this method in determining the influence of stereotypes upon people's expressions of attitude.[25] Statements of definite "class" or "party" significance were submitted to comparable groups and to the same groups after an interval of time. The statements were first submitted *without* any characterization such as "conservative," "radical," "Communist," "Fascist," etc., and were then submitted *with* these labels attached to the statements. The resulting changes in the degree to which different groups accepted or rejected the statements indicated the degree to which these groups were influenced in their attitude toward an issue by the presence or absence of a stereotyped label. Menefee used a somewhat similar technique in studying the effect of emotional as compared with relatively non-emotional propaganda.[26]

A somewhat different and more highly controlled type of experiment to determine people's attitudes and the factors that influence them is represented by such studies as that of Horowitz,[27] who studied the development of attitudes in children toward the Negro by subjecting them to photographs of play groups, eating groups, etc., some of which had Negroes present and others not. Choices by a child of the groups he preferred to join indicated the presence or absence of prejudice against the Negro.

A very large literature of this general character reporting studies of changing attitudes under more or less controlled conditions has already appeared. These studies will probably increase and improve in quality.

Studies of the type described above, as well as those of a controlled laboratory sort, are dependent upon the use of statistical methods in recording and interpreting the findings. We are interested in observing the differences in behavior under different circumstances. But when are ob-

[25] S. Menefee, "The Effect of Stereotyped Words on Political Judgments," *Amer. Sociol. Rev.*, 1:614-21, 1936.

[26] S. Menefee, "Propaganda and Opinions on Foreign Policy," *Jour. of Soc. Psych.*, 11:393-404, 1940. See also the intensive study (forthcoming) of the effect of a presidential campaign in Erie County, Ohio, 1940, by Paul Lazarsfeld. See also p. 365 below, note 12.

[27] E. L. Horowitz, "The Development of Attitude Towards the Negro," *Archives for Psych.*, No. 194, 1936. See Chap. VIII of the present text for a summary of Horowitz's methods.

served differences attributable to chance, and when are they almost certainly due to the conditions of the experiment? These are crucial questions in any experiment, and they can be answered only on the basis of the known laws of sampling, probability, and error. These are subjects which occupy a central place in any treatise on statistical methods, and although we shall make some references to those subjects in later chapters, a full discussion of the statistical methods indispensable to scientific experimentation in any field must be left to the treatises on that subject. The usefulness of simple statistical analysis in what Chapin[28] has called *ex post facto* experimental studies should, however, receive some attention in the present connection. The approach is described by Chapin in part as follows:

"From diverse experiments with the experimental method in education, psychology, and sociology, a pattern of practicable procedure has begun to emerge. It is our opinion that this pattern of procedure supplies the outlines of a long desired design for social experiments. To avoid misunderstanding of terms, let us state at the outset that by experiments we do not mean such trial and error efforts as the NRA, the AAA, the TVA, or similar gigantic social-economic reforms which are without controlled variables or adequate devices to measure their effects. We mean by experiment the observation of the changing relationship in an interval of time between two variables, meanwhile holding constant or controlling several other variables, which if uncontrolled might themselves explain or cause the effects. Suffice it to say that we shall attempt to describe briefly the efforts made to adapt the experimental method of physical science to the study of cause and effect or functional relationships in the social field. . .

"This experiment[29] was based upon the high school rec-

[28] F. S. Chapin, "A Study of Social Adjustment Using the Technique of Analysis by Selective Control," *Soc. Forces*, 18:487, 1940. See also, by the same author, "Design for Social Experiments," *Amer. Sociol. Rev.*, 3:786-800, 1938.

[29] Helen F. Christiansen, "The Relation of School Progress, Measured in Terms of the Total Amount of School Attendance or Course Completion, to Subsequent Economic Adjustment." M.A. thesis, June, 1928, Univ. of Minnesota Library. See also a similar study by N. G. Mandel, "A Controlled Analysis of the Relationship of Boy Scout Tenure and Participation to Community Adjustment," M.A. thesis, Univ. of Minnesota Library, July, 1938.

ords and community experiences of 2127 boys and girls who left four St. Paul high schools in the school year of 1926, as graduates, or after having completed from one to three years of their high school course. There were only four St. Paul high schools at this date so that the experiment concerns the whole universe and not a sample. The year 1926 was taken because it was the earliest year for which comparable records on a large number of students were available. Moreover, since the follow-up was to the year 1935, there was thus a period of nine years in which these individuals could work out economic adjustments.

"The working hypothesis of this study was: a greater degree of progress in high school leads to a correspondingly higher degree of economic adjustment in the community. Thus, the two variable factors whose relationship is to be measured are ones that lie at the heart of educational policy. Do we not spend millions of dollars annually to support high school education on the assumption that this hypothesis is a true statement of social relationship? Of course, there are other purposes of high school education besides economic adjustment, but certainly economic adjustment is one of the most practical considerations.

"Having set up the working hypothesis of the experiment, it next becomes necessary to secure measures of the two variables to be observed. The independent variable, school progress, was measured by the number of years of the high school course completed when the student left school in 1926. Of the total 2127 boys and girls, 1130 graduated from high school in 1926 after completing four years and 997 dropped out in 1926 after having been in high school for the regular one or two or three years of the course. The measure of economic adjustment selected for the dependent variable was the percentage of shifts in jobs from 1926 to 1935 that involved no change in salary or an increase in salary as contrasted to the percentage of shifts that involved decrease in salary.

"Now it is perfectly obvious that these are extremely crude measures. Factors of age difference as between those who left at the end of the freshman high school year and those who remained to graduate might affect economic adjustment. Sex

differences are often significant. Boys or girls from homes of higher status would have an advantage in gaining and holding employment not possessed by children from homes of lower status. Differences in the nationality of the parents would influence the chances of getting a job. The neighborhood of the home from which the boy or girl came might be a factor in economic adjustment. The intelligence or mental ability of the different individuals would exert its influence upon securing a job, holding the job, and upon promotion in rank and salary on the job. This network of factors is also one of the interrelationship. An American boy whose father was a successful professional man living in a restricted neighborhood would have several factors combined in his favor in comparison with the Jewish boy whose father was a clerk and lived in a deteriorated neighborhood. Since every one of these variable factors are recognized by sensible people as influencing the course of individual economic progress, the way to obviate their disturbing influence is to control them. Here, therefore, we come again to our crucial problem of controls. . .

"Each of these six factors, chronological age, sex, nationality of parents, father's occupation, neighborhood status, and mental ability (by computing high school marks because the measurement by I.Q. was not available) was controlled. . .

"It took a full year of systematic work in home visits and interviewing to trace the 1130 graduates of 1926, and the 997 drop-outs of 1926, to their status of 1935. In this process, there was a shrinkage of 933 individuals in the total. Of this number lost, 21 were deceased, 42 had moved out of town, 575 could not be traced in the follow-up, and 295 had records so incomplete as to make comparison worthless. Thus, of the original 2127, there were located a group of 671 graduates and a group of 523 drop-outs. . .

"The process of gaining control began with the selection from the control group of a child who was then matched with another child from the experimental group for sex and nationality of parents. This reduced the two groups to smaller groups with identical proportions in sex division and in the distribution of parental nationality. At this point the control of factors by identity through individual matching had to be supplanted by control through the correspondence of

frequency distributions on each factor. The reason for this change was that the conditions of individual identity on a factor by matching eliminated so many cases that the sample dwindled in size at an alarming rate after each new control was set.

"The correspondence of frequency distributions on a given variable factor is a far less rigorous control of this variable factor than is identity by matching individual with individual, but it is probably the most frequently used method of control in contemporary experimental studies. Its selection and application on grounds of reducing the attenuation of the sample is justified when the results of the experiment are significant. When the results are not significant and no other explanation is reasonable, then it is necessary to resort to the more expensive process of individual matching, but this means repetition of the experiment, or at least increasing the number of cases observed.

"Setting the six controls reduced the final sample to a total of only 290 cases, 145 in the control group and 145 in the experimental group, a decline of 86.4 percent from the original group of 2127 students! This is the price of observation under conditions of control. The longer the list of controls and the more vigorous their method of application, the smaller the final sample. At this point, some statisticians may say that we end up with a sample too small to be representative of conditions in a large original group, but let me remind them that an experiment is designed to obtain a homogeneous and 'pure' sample. In fact, our sample has been purged of the very factors that made for heterogeneity in the original group and whose presence obscured the real relationship between the factors we set out to study. To discover the *real* relationship between a magnet and iron, we must have 'pure' iron and not iron ore that is complicated by the presence of other minerals and metals, which it would be if 'representative' of the original ore. Homogeneity, not representativeness, is *the* essential condition to the discovery by experiment of a real relationship between two factors. Consequently, if the present experiment shows that there *is* a relationship between the amount of high school education and the degree of subsequent economic adjustment, this re-

lationship is more likely to be a *real* relationship than is the case under conditions in which the social situation is complicated by uncontrolled factors. Within certain limits (the limits of the given experiment), the homogeneity of the two sub-samples is more important than their representativeness of variable factors that originally obscured the relationship.

"Finally, if we turn now to the differences in economic adjustment of the control group of drop-out students and the experimental group of graduates we find that 88.7 percent of the graduates experienced no changes in salary or had increases in salary from 1926 to 1935, whereas 83.4 percent of the drop-outs reported increases or no changes in salary from 1926 to 1935. Putting it the other way, only 11.3 percent of the graduates suffered salary decreases in this period, whereas 16.6 percent of the drop-outs suffered salary decreases. This is a small difference of 5.3 percent.

"When the length of high school education before drop-out is analyzed, we find that 74.1 percent who left school in 1926 at the end of one year of high school had salary increases or no changes in salary during the period 1926–1935; and of those who ended two years of high school, 85.1 percent were adjusted economically; and 89.6 percent of those who ended three years of high school were adjusted. Thus, in general, the longer the period of high school education, the higher the percentage of adjustment in the economic terms used as a criterion. *None of these differences are statistically significant as single differences.* The important point is that they are consistent and in the same direction.[30] It is a matter of opinion whether small differences that are corroboratory and in the same direction are as important as differences that are large and statistically significant. Our opinion is unequivocally that small differences in the same direction may be as important as one large difference that is statistically significant. There are two reasons for this opinion. First, social phenomena are complex and not likely to show large differences because of the configurational character of the social situation. The separate factors in a social situation are usu-

[30] For an important discussion of the statistical aspects of this point, see C. C. Peters, "Note on a Misconception of Statistical Significance," *Amer. Jour. Sociol.*, 38:231-36, 1933.

ally functionally related. Second, the conventional tests of the significance of sampling are based upon the theory of random samples, and in experimental work, as we have attempted to show, it is the terminal homogeneity and purity in the sample, rather than initial representativeness of heterogeneity, that is important in demonstrating the real relationship between two variable factors.

"Thus [this] experiment, in spite of the crudity of measuring devices used, was nevertheless a successful experiment if the argument in favor of corroboratory small differences is accepted, because these small differences in salary changes, percentage voting, average number of interests in activities, average number of additional years education, and occupational class attained, all support the hypothesis by showing evidence of better adjustment with every increase in the amount of high school education. Doubtless there will be those who will observe that the experiment has merely proved that which we expected to be true. But justification by expectation and wishful thinking is quite a different thing from corroboration by factual evidence. On the other hand, there may be critical persons who will observe that the results of this experiment are almost too good to be true. To all such, we may reply that the evidence is available for critical appraisal in the thesis, and that in any event, we do not claim one successful experiment proves a relationship to be true. The experiment needs to be repeated and the results verified before the *principle* is established on factual evidence rather than on wishful opinion. . .

"In this experiment, we have found that control of such variable factors as age, sex, nationality of parents, occupation of father, neighborhood status, and high school marks, is sufficiently adequate control of the social situation to lead to definite scientific results. What would have happened if we had exercised a more rigorous control by the device of identity secured through individual pairing for every variable factor instead of relying on the rough control of correspondence between distributions in the case of the four factors of age, father's occupation, neighborhood status and high school marks? . . .

"The answer to this question [has been found] by actually

pairing against each individual in the control group of drop-outs, another individual from the experimental group of graduates of the same sex, identical chronological age, the same parental nationality, the same father's occupation, the same neighborhood status and identical high school marks. This process of more exact control reduced the total from 290 to only 46, of which 23 were in the control group of drop-outs, and 23 in the experimental group of graduates. When the economic adjustment of these two contrasting groups is examined, the following results emerge: 92 percent of the experimental group of graduates had salary increases or no change in salary from 1926 to 1935, whereas only 58 percent of the control group of drop-outs experienced salary increases or no salary change for the period. This is a difference of 34 percent in favor of the graduates. In comparison, the difference was only 5.3 percent for subgroups of the 290 cases under less rigorous controls. Thus, the labor of more exact control was amply repaid because it demonstrated in a very decisive manner and in terms of a difference statistically significant, the superior economic adjustment of the experimental group of graduates. In this case, the real relationship was more clearly demonstrated by more rigorous controls." [31]

The above example will give some idea of the importance of statistical method in social experiments. Not only is this method essential in recording and measuring the variables in experiments, but it provides a technique for determining the relationships of factors and the *degree of contingency* between them. This is all that even the comparatively exact sciences can do. As Pearson has pointed out: "That the universe is a sum of phenomena, some of which are more, others less closely contingent on each other is the conception wider than that of causality, which we may at the present time draw from our widening experience. . . No phenomena are causal; all phenomena are contingent, and the problem before us is to measure the degree of this contingency, which . . . lies between the zero of independence and the unity of causation." [32] Therefore, concludes Mills, "Association and

[31] F. S. Chapin, "Design for Social Experiments," *op. cit.*, pp. 786, 789-94.
[32] Karl Pearson, *The Grammar of Science*, (Third Edition, A. & C. Black, London, 1911), pp. 173-74.

correlation are the terms which replace causation, as probability and approximation replace the concept of certainty." [33]

By the method of sampling, averages, frequency distribution, and selective control as described above, the statistical method disposes of the apparently insuperable difficulties of the enormous number and variability of social phenomena. By the method of correlation it provides a technique of measuring the degree of contingency between social phenomena. The importance of these techniques will be elaborated in later chapters. It is to the statistical method that we must look in the social sciences for an escape from the seemingly insurmountable obstacles to precise methods of experimentation in sociology.

While social experimentation will probably always lack the rigor of the experiments of the physical laboratory as far as artificial control of the conditions is concerned, the same result will be achieved through refined techniques of measuring and allowing for the influence of these conditions. In the practical applications of the physical sciences such calculations and allowances are always made. Such factors as altitude, temperature, etc., are variables always present, and which have to be measured and allowed for, before the "scientific laws" describing the situation are valid. The technique for holding constant certain factors in order to study the influence of some other factor without artificial manipulation of the actual social conditions involved has already been developed to some degree in the statistical device of partial correlation. Social experimentation, therefore, instead of aiming to isolate the experiment by an actual elimination of all the complicating physical and social factors will instead merely allow for them through the methods of measurement provided by statistics.

Q. SUMMARY AND CONCLUSIONS

The purpose of this and the preceding chapter has been, first, to emphasize the need of a scientific approach to the study of group behavior, and especially the need of objectivity in observation and recording. The unaided senses are

[33] "On Measurement in Economics," ch. 2 of *The Trend in Economics*, Edited by R. G. Tugwell, (Knopf, 1924).

subject to very grave shortcomings, (*a*) because of their inherent limitations, and (*b*) because of the influences of bias and prejudice. The observation of social phenomena is especially subject to these influences. Moreover, the requirement of science that the essential facts of a situation shall appear the same to all competent observers has led to the development of instruments and devices for objectifying, standardizing, and accurately measuring observations. But so far, these devices and methods have been developed almost exclusively in the physical sciences. There the experimental method has been our chief reliance.

In the observation of social phenomena there exist certain grave obstacles to the use of the experimental method, and consequently to the development of refined methods of observation, measurement and causal inference. These obstacles spring, first, from the opposition of society to any active interference with the determining conditions of a social problem. Questions of human rights, freedom, morals, and other subjective, sentimental, and emotional considerations are involved. Second, there is the very difficult if not impossible matter of securing isolation or proper control of all the variables in a social experiment because (*a*) it usually has to be performed in a social medium, and (*b*) all the variables may not be suspected.

On account of these obstacles in the social sciences to the type of experimentation carried on in the physical sciences, the most promising and productive form of social experimentation has become that provided by statistical techniques. Such techniques permit us to measure the influence of the known variables, and thus to allow for their influence in any given problem. This is equivalent to holding these variables constant, while the operation of some one factor is observed. Thus the essential conditions of a scientific experiment are achieved without the actual physical or social manipulation of the conditions that are significant to the experiment. Through the statistical methods of sampling, averaging, and correlation, generalization and causal inference become possible. Through such techniques also, tools of social measurement and prediction have developed in such forms as scoring schedules, business barometers, business and social indices,

and actuarial tables. In these devices for forecasting the social weather we have the counterparts of the barometer, the thermometer, and the other instruments of physical science.

While these considerations clearly indicate the possibility of scientific method in the field of social phenomena, it should be frankly recognized that as yet many of the necessary techniques are highly undeveloped. Furthermore, we have not fully realized the need in our social organization of using to the fullest extent the techniques now available. Important groups of social facts are either not collected at all, or exist in the crudest form. The increasing adequacy of the United States Census and of the statistical records kept by other governmental agencies, such as those of the Department of Commerce and the Public Health Service, represents hopeful developments for the future. States and cities supplement these sources through censuses and record-keeping of their own, but again the incompleteness, incomparability, and general inadequacy of these records greatly limit their usefulness for scientific and experimental purposes. However, efforts are now being made to correct these as governmental agencies seek statistical advisers. The long time usually required for social experiments makes it of the greatest importance that social statistics be standardized and that their comparability in various geographic areas and over periods of time be maintained. Finally, the lack of development in statistical method itself imposes certain limitations on its use in more refined scientific generalizations. But up to the present our technique of handling data already collected far surpasses in refinement the data themselves and the methods of securing them. A comprehensive program of *careful and standardized recording of data* is the greatest immediate need for the beginning of a scientific understanding and control of society.[34]

There are many able and hard-working people who are impatient with the painstaking methods of science and the seemingly futile recording of detailed data over long periods of time. These workers overlook the fact that often the chief reason for the difficulties of solving some of their problems is

[34] The expansion and improvement of official statistics will be reviewed in Chap. 11. See also W. F. Ogburn, "Statistical Trends," *Jour. Amer. Statis. Asso. Proceedings of the Centenary Celebration 1839–1939*, 35:252-60, 1940.

that we have not yet formulated the principles to apply to their problems. These principles can be arrived at only through collection and manipulation of data by constantly more refined methods.

In this discussion of the promise of scientific method in the social sciences there is no desire to deify science. The limitations of scientific method, and the tentative nature of the conclusions reached by it should be clearly recognized. In some quarters [35] the thesis is advanced that science is to the great masses merely another folklore to which we look for explanation and guidance. This new folklore has its own priesthood, namely the scientists. The belief of the common man in science is in all important respects of the same kind as his belief in other folklore. He can no more investigate for himself the basis of his belief in science than he can investigate the basis of his belief in other lore. He accepts both on faith, if at all. Thus, his belief in the rotundity of the earth is essentially similar to his belief in the revelations of Moses. He accepts both on authority. It may be objected that the theologian quotes authority as a necessity and enjoins belief as a duty, while the scientist quotes authority purely as a matter of convenience and can meet disbelief with demonstration. That is, he who doubts the findings of the scientist may after a long period of training himself become a scientist and thus may become privy to the sense experience on which scientific belief rests. But in the same way a doubter of theological doctrines may himself become a theologian and examine the basis of these doctrines, whether from ancient records or direct revelation. The means of verification of either type of lore is beyond the reach of the great majority of men, and is fundamentally of the same nature. Both rest, in the last analysis, upon the similarity of the responses of a number of individuals to given phenomena.

To many people this analogy is considered as proof that science is no more entitled to our respect or reliance than any other folklore. This view is further supported by the fact that the history of science even in its most highly developed fields is marked by a rapid discarding of old postulates and concepts for new. From these considerations the conclusion

[35] See C. E. Ayres, *Science the False Messiah,* (Bobbs-Merrill, 1927).

is drawn that, since nothing is final or absolute, and since all theories must be taken on faith, one folklore is as good as another. But this is clearly an erroneous conclusion. There are three principal tests of the validity of a method or of a conclusion. The first is the agreement of competent investigators; the second is the agreement and congruity of the observations and conclusions with each other; the third test of validity lies in the utility of a given method in solving our problems. In all three respects, the methods and conclusions of science have proven, not their infallibility, but their superiority. It is by these criteria that the relative value of folklore must be determined. It is by these criteria that science must make its appeal. And it is by these criteria that the true scientist always regards his conclusions as tentative, and to be discarded whenever a new conclusion or a new method promises to meet the test of utility more adequately.

Whatever view one may take regarding the possibility of the development of exact social sciences as set forth in these chapters, all will agree to the immediate and practical value of the collection of certain social data for guidance in the everyday problems of administration of current affairs. The principle of ordinary bookkeeping, at least, will need no defense. To the technique of collecting such data as is generally conceded to be essential, we now turn our attention.

As for the more ultimate aim of exact social sciences, it may be suggested that in the current skepticism on the subject of the actual usefulness of scientific method in the solution of concrete social problems, science is at least entitled to the benefit of the doubt. As Chase has said: "So far as we can see, the future . . . lies with the man of science — the social scientist, the engineer. For upwards of one hundred thousand years, the mystics, the medicine men, the orators, the spellbinders . . . the personally crafty, the dealers in . . . rationalized sophistry, have imposed upon others their shoddy inferences, conclusions and even controls. It is difficult to see how the man of science can do worse." [36]

[36] Stuart Chase, *The Tragedy of Waste,* (Macmillan, 1930), p. 279.

R. SUGGESTIONS FOR FURTHER STUDY

1. For a more general discussion of the problems of objective observation of social phenomena, see G. A. Lundberg, *Foundations of Sociology,* pp. 17-22, ("The Visibility and Objectivity of Social Phenomena"). See also, A. F. Bentley, *Behavior Knowledge Fact,* (Principia Press, Inc., Bloomington, Ind., 1935), Chap. 21 ("The Visibility of the Social").

2. For a comprehensive digest of the large literature of concrete studies in social psychology with special reference to experimental studies, see G. & L. Murphy and T. Newcomb, *Experimental Social Psychology* (Rev. Ed., Harper, 1937).

3. Excellent brief collateral reading for the sections on experimental methods will be found in the three experiments (one of which is summarized in the text) described in F. S. Chapin's "Design for Social Experiments," *Amer. Sociol. Rev.,* 3:786-800, 1938. Further description of this procedure may be found in the following subsequent articles by the same author: "A Study of Social Adjustment Using the Technique of Analysis by Selective Control," *Social Forces,* 18:476-87, 1940; "An Experiment on the Social Effects of Good Housing," *Amer. Sociol. Rev.,* 5:868-79, 1940.

4. For a profound consideration of the logic of scientific inquiry, see John Dewey, *Logic: The Theory of Inquiry,* (Holt, 1938), Part IV.

5. Herbert Spencer's *The Study of Sociology,* (Appleton, 1903), remains a classic on the subject of the present chapter. On the different types of bias likely to be present in sociological study, see especially ch. 8-12.

CHAPTER III

TERMINOLOGY, UNITS, AND CLASSIFICATION

The rough working definitions of the dictionary serve to throw some light on the meaning of isolated words, but it is when the words are actually used in assertions that their inherent indefiniteness leads to doubt, and to this is due nine-tenths of the wrangling that goes on in all the Council chambers of the country. — *F. W. Westaway.*[1]

The preceding chapters have emphasized that one of the principal requirements of a scientific fact is that it shall appear the same to all competent observers. This requirement makes it necessary (1) that the conditions of observation be standardized, and (2) that a standardized set of symbols and terminologies be employed to describe our sense experiences. The first of these conditions was treated in the chapter immediately preceding. The present chapter considers the second requirement.

A. THE IMPORTANCE OF DEFINITE TERMINOLOGY IN SCIENCE

The futility of much argument and discussion resulting from a poor definition of the terms employed could be illustrated at great length. A typical example from Professor James will be sufficient here.

Some years ago, being with a camping party in the mountains, I returned from a solitary ramble to find everyone engaged in a ferocious metaphysical dispute. The *corpus* of the dispute was a squirrel — a live squirrel supposed to be clinging to one side of a tree-trunk; while over against the tree's opposite side a human being was imagined to stand. This human witness tries to get sight of the squirrel by moving rapidly round the tree, but no matter how fast he goes, the squirrel moves as fast in the opposite direction, and always keeps the tree between himself and the man, so that never a glimpse of him is caught. The resultant metaphysical problem now is this: Does the man go round the squirrel or not?[2]

Obviously, as Professor James soon pointed out, this whole dispute hinged on the meaning of the term *go round*. If we

[1] *Scientific Method*, (Blackie, London, 1912), p. 8.
[2] William James, *Pragmatism*, (Longmans, 1907), p. 43.

mean by *going round* to occupy successively positions, say, east, south, west, and north, of the object in question, then certainly the man goes round the squirrel. If, however, we mean by *going round* to occupy successively positions in front of the squirrel, to his right, behind him, and to his left and finally in front of him again, then it is equally clear that the man does not go round the squirrel, for the squirrel in the problem is constantly changing his position so as to keep it constantly relative to the man. Now our point is that the great majority of disputes have at their basis some such misunderstanding or conflicting use of terms.

The importance of language[3] as an instrument of adjustment among human beings is usually emphasized in all general treatises on sociology and psychology. It is pointed out, for example, that a map is a highly valuable symbolic representation, provided its pattern, order, and sequences correspond to the actual terrain over which we must travel. It is a corresponding handicap, if it fails to indicate where the rivers and the mountains are, if it confuses the order in which they occur, or otherwise fails to correspond to the conditions to which we must adjust. Sometimes a local map which may be adequate for most kinds of travel within its own borders turns out to be very misleading in the light of a larger perspective. In the same way the verbal systems and orientations of a primitive, primary-group society may turn out to be grossly inadequate in a national or world society of secondary group relationships.

We are not here primarily concerned with the vast amount of misunderstandings among people in daily conversation on account of the inadequacy of their language as a means of communication. We are interested mainly in the same problem as it affects the social sciences. As an illustration of the type of impasse into which an inadequate symbolic system is likely to force us, we may refer to Zeno's well-known paradox

[3] For a good popular treatment of the subject, see Stuart Chase, *The Tyranny of Words*, (Harcourt, 1938). An excellent brief treatment is Leonard Bloomfield, *Linguistic Aspects of Science*, Encyclopedia of Unified Science, Vol. 1, No. 4, (Univ. of Chicago Press, 1939). See also, E. A. Esper, "Language," Chap. 11 in *Handbook of Social Psychology*, Edited by C. Murchison, (Clark Univ. Press, 1935), p. 456.

of Achilles and the tortoise. If Achilles allowed the tortoise a head start in a race, it was suggested, the former could never overtake the latter. Whatever common sense and an actual trial might show, the language and the logic in terms of which the paradox was set up left no escape from the strange conclusion. The tortoise is allowed one hundred yards start. Although Achilles runs ten times as fast as the tortoise, the latter will obviously have gone an additional ten yards by the time Achilles has run the first hundred yards. When Achilles has run this additional ten yards, the tortoise will still be one yard ahead. By this process of verbalization about the race it is obvious that, while Achilles is constantly getting nearer to the tortoise, he can never quite overtake it.

If instead of the above method of symbolizing this event we resort to a simple algebraic or geometric statement[4] of a type unknown to the Greeks, we come to a very different conclusion. We find by the latter methods (under given assumptions) that Achilles definitely overtakes the tortoise at the end of 111.11 seconds. What is more important, this result also corresponds to what happens in an actual experiment. This verifiable result of the latter method suggests that perhaps it is preferable as a symbolic device for such mundane matters as, for example, planning the dispatch of trains. The former method may be more advantageous in "creating" thrilling situations in imaginative literature, just as the trips of rocket-ships to the moon in current cartoons and movies have unquestioned recreational value. They may be regarded as quite harmless unless someone actually attempts to take off, with present equipment, on such a journey. We are interested here in symbolic systems which will be reliable

[4] See the geometric demonstration in L. Hogben, *Mathematics for the Million,* (Norton, 1937), p. 89. In practical adjustments to our immediate environment we set aside symbolic representations which do not correspond to the observed facts. But in our adjustments to secondary group situations with which we have never had direct personal experience, we rely entirely upon symbols, and if they do not faithfully represent the concrete adjustments to be made, the results are disastrous in the same way that it would be disastrous in certain circumstances to rely on the reasoning involved in the story of Achilles and the tortoise. Many of the pronouncements of social philosophers are dangerous in the same way. The symbols, the reasoning, and the conclusions seem irrefragable on paper, but they simply do not correspond with what occurs in actual social relations. Many proverbs are of this character. They represent not statements of facts, as is commonly assumed, but somebody's wishful thinking as to what *ought* to occur.

charts to adjustments beyond the universe of the symbols themselves.

The symbols man uses at first to represent the world are oral, and very little scientific development is possible so long as this is true. Many of the things that scientists wish to communicate simply cannot be adequately transmitted through oral language. The structure of the idea which scientists must communicate is so complex that it cannot be matched in a succession of acoustic stimuli. Consequently, the language of science must increasingly consist of written graphic symbols, which provide an enduring instead of an immediately vanishing stimulus, and which offer possibilities of arrangement (tabulation, etc.) that cannot be communicated in oral language. In devising written symbols, the tendency in science is to develop, as soon as possible, special symbols representing highly abstract, standardized, and ordered responses called quantitative terms instead of the emotionally charged words employed in everyday common-sense communication. The indefiniteness of meaning of such symbols as "many," "much," "more," "few," "less," "least," as well as avowedly "qualitative" terms such as "good," "bad," "symmetrical," "level," "heavy," "light," etc., led to the invention of number systems and the symbolic division of various subject-matters into *units* and scales. These are highly standardized and readily verifiable response categories.

So much has science been dependent upon this development that we find the advancement of science largely contingent upon the development of mathematics. It is on the symbolic uniformity of mathematics that we rely for further power to describe the uniformities in the behavior of phenomena. The description of these uniformities constitutes the essence of science.

B. CONFUSION OF TERMINOLOGY IN THE SOCIAL SCIENCES

In the description of social phenomena very little use has yet been made of the language of science, namely, mathematics. As in the early stages of every science, we still employ in the social studies the vague and indefinite terminology of common parlance drawn from the folklore (common-sense experience) of the past. That terminology, consisting

as it largely does of the concepts of primitive mysticism, is not well adapted to the uses of science. But the meaning of words, being the growth of centuries, is not easily abandoned or changed. Even in the comparatively well developed physical sciences, there is still a large vocabulary of common-speech designations which are retained as traditional units. Thus we have in physics such terms as *horse-power, foot, candle-power,* etc. The difficulties inherent in such terms have been obviated by assigning to these traditional words mathematical definitions that are entirely arbitrary. A horse-power, for example, originally meant the amount of energy which a horse is capable of exerting. Clearly, this amount would vary within wide limits according to the size and physical conditions of the horse. But the scientific definition of the term as the amount of power required to lift 33,000 pounds one foot high in one minute at sea level permits a very much more accurate designation of the idea. One of the principal reasons for the comparative indefiniteness of the social sciences is that in these studies we still employ the terminology drawn from common-sense experience without having assigned to these terms a meaning approaching in definiteness the physicist's definition of *horse-power.*

The results of this lack of a standardized terminology in the social sciences are apparent in nearly all controversies and discussions in these fields. Such common terms as *group, society, community* and *public* are frequently used interchangeably and with different meanings by different writers. The result is much fruitless argument over the meaning of such terms as *public opinion, social mind, community, consciousness,* etc., and a great vagueness as to which of several common meanings of a word should apply in a given case. In the social sciences, most of the important terms have several meanings. Usually the meaning is dual — one being the meaning of the term as employed in common parlance and the other a broader or a more limited meaning which, by common consent, has been adopted by the students in a particular field. Thus *culture* is used in common speech to indicate the so-called refinements and etiquette of polite society, usually including some familiarity with *belles lettres* and other fine arts. In anthropology and sociology, *culture* usu-

ally includes the entire social accumulation of behavior, customs and institutions, but there is much vagueness and difference of opinion even in these professions as to the exact meaning of the term. A college student once questioned the author's use of the term *behavior* to include both *good* and *bad* behavior. This student had always thought of the term as applying only to socially approved conduct. Likewise, the confusion in the meaning of the word *primitive* has caused the proposal of the term *pre-literate* for societies which have not yet developed a written literature. Again, in economics there is a wide disparity between the popular and the economist's definition of *production, consumption,* etc.

In 1905 Small listed forty-eight terms and concepts which at that time, in his opinion, constituted the terminology of sociology.[5] In 1931 Eubank, after an extensive survey of the literature of the field, compiled a list of 332 such terms and concepts.[6] Commenting on his list, Professor Eubank says:

> The very extent of such a list as this is confusing; but it is indicative of an underlying confusion that is still more significant. It reveals strikingly how far the sociologists are from agreement upon the very terminology itself. The youth of this science is nowhere more clearly shown than in the indefiniteness and lack of uniformity of its vocabulary. Different writers use the same term in different senses. Conversely, the same idea is labeled differently by different writers.[7]

As an illustration of this confusion, Eubank has tabulated the lists of major concepts employed by each of ten leading American sociologists, as found in eight general texts. He finds a total of 146 items. Only 63 out of the 146 are found in more than one list; not one appears in all eight lists; only one term appears in as many as seven lists; only two others are found in six lists; and only eight other terms appear in five of the lists. Professor Eubank summarizes the matter as follows:

> Expressed in a sentence, this means that out of eight volumes of sociological theory, selected because their authors are men whose recognized standing in American Sociology entitles them

[5] Albion W. Small, *General Sociology,* (University of Chicago Press, 1905), Chap. 29.
[6] E. E. Eubank, *The Concepts of Sociology,* (Heath, 1932), Chap. 4.
[7] *Ibid.,* p. 43.

to be regarded in an especial way as representative of the field, not a single concept occurs upon which they are agreed as a main heading for the materials which they have singled out as most important to discuss; and over 55 per cent of the concepts used as main headings (a total of 83 out of 146) are found in not more than one of the entire eight tables of contents.[8]

The incomparability, and hence the diminished scientific utility, of much current social data, because of the lack of objective definition of units, is strikingly illustrated in the field of social work. McMillen reports the results of a study in this field which revealed "a complete absence of uniformity among agencies in their classification of major care and minor care cases. . . Moreover, the data afforded no evidence of uniform policy by which minor care cases could be differentiated from the miscellaneous services not counted as cases. The responses showed unmistakably that a contact which might be counted as a minor care case in one agency would in another be excluded from the service count."[9] A further similar check on 25 agencies reputed to have high standards of work "revealed a lack of uniformity of practice very little less marked than that found in the entire group." The movement during the last decade for standardized reporting of social statistics (see pp. 402-406) has undoubtedly greatly improved this situation. It is obvious that intelligent discussion, not to mention statistical studies, depend on a more rigorous definition of our terms.

C. THE CRITERIA OF SCIENTIFIC TERMINOLOGY

It is clear that no field of inquiry can be reduced to a very definite science as long as its terminology and concepts are thus badly defined. The need for terminology understood and used in the same way by all the students in the field is, therefore, a great present need in all the social sciences, and perhaps especially in sociology. This gives rise to the question how such a terminology is to be developed. The ulti-

[8] *Ibid.*, p. 46.

[9] A. W. McMillen, *Measurement in Social Work*, (Univ. of Chicago Press, 1930), p. 48. See also A. W. McMillen and Helen R. Jeter, "Statistical Terminology in the Family Welfare Field," *Social Service Review*, 2:357-84.

mate test of the value of a particular term or concept must be its usefulness for the purpose at hand, namely, the description of behavior so objectively as to be subject to corroboration by other persons, usually by reference to standards external to the individual organism. Terms and categories should be regarded merely as working tools, and not as entities in which inhere magical properties. They should, therefore, be modified or abandoned as they serve, or fail to serve, our purpose. But at the same time, the requirements of science make necessary a certain uniformity. This uniformity can be secured only by general agreement between the workers in a field to use certain symbols to designate certain facts or relationships. Thus the American Statistical Association once had a Committee on Terminology to secure uniformity in the use of Greek letters and other terms as statistical symbols. In 1938 a Committee on Conceptual Integration was formed in the American Sociological Society to consider ways and means of remedying the confusion reviewed above. Similar action is badly needed in other social sciences.

The criteria by which a term is to be accepted or rejected must again be determined ultimately by the utility of the results. Professor Eubank suggests four tests of a true scientific concept: [10]

(1) Is the term reasonably precise? Does it convey an exact and clear-cut meaning?

(2) Does the term convey only one final idea?

(3) Is the term perfectly general; that is, always employed in the same sense wherever it is used?

(4) Is the idea fundamental to its particular field; that is, essential for its inclusive interpretation?

While several of these criteria are themselves somewhat subjective, they may serve as tentative guides to the formulation of a useful terminology. Uniformity in the use of terms thus arrived at can be secured only by common agreement. This is important. On the other hand, if the specific meaning of a term is objectively defined, and if it denotes a significant behavior-segment or operation, it matters little what symbol we agree upon.

[10] *Op. cit.*, p. 398.

D. THE NATURE OF OPERATIONAL DEFINITIONS IN SCIENCE

Perhaps the most elementary way of defining a word is to point to an example of it or to go through a motion and simultaneously say the word used to designate that motion (or set of motions). That is, we teach a child the meaning of the word "book" by pointing to one or several books in succession; we leap into the air and say, "jump"; and so forth. The ease and clearness with which we can communicate the meaning of a word in a given case depends here upon the convenience with which such motions can be performed. The clearness with which we can communicate the *general* or *class meaning* of these words depends on the facility with which the motion in a given case will be recognized as similar to these objects or motions in the cases of other books and other jumps.

The ease with which different words can be defined in this way, and consequently the uniformity of meaning which different users of a word attach to it, varies greatly. The difficulty increases as we desire to communicate the meaning of abstract words, designating complicated relationships between events and things. Such words as electron, acceleration, honor, status, cooperation, etc., are of the latter type. The problem of definition of scientific concepts is largely of this nature.

We have already pointed out what physicists have done with such terms as *horse-power* when they have taken over such a folk word as a designation for some aspects of the phenomena with which they work. As we have seen, physicists invent instruments which behave in certain ways — which "register" — when applied in certain ways. Clocks, thermometers, ammeters, etc., are of this character. Scientists are today content to define such concepts as *time, temperature, electricity,* as *that which* is registered or measured by these instruments. Questions about the "ultimate nature" of these phenomena are regarded as obsolete and unscientific. Scientists agree to designate each degree or kind of behavior which their instruments indicate by specific words or other symbols. These words so defined may then be used to build up more complicated words, the definition of which is, however, always reducible to the readings of standardized instruments.

Thus, velocity is a rate of motion, acceleration is a rate of change of velocity. Even such a word as "electron," which is popularly thought of, perhaps, as a small billiard ball, is now defined by the physicist by means of a differential equation. Such definitions are called operational definitions.

There is nothing mysterious or magical about the word "operational." It is merely a word we use to designate a type of communication which "gets across" with high reliability. Thus, a recipe for a chocolate cake may be regarded as an *operational definition* of such a cake. Now a recipe is, of course, only a group of words, and, as such, is no different from any other group of words constituting a definition. Strictly speaking, we might contend that it is the *actual performance of baking a cake* which constitutes its definition. To the extent that the words of the recipe designate accurately all the steps in the performance, however, we call the recipe an operational definition of the cake in question. The possibility of writing such a recipe obviously depends upon the prior existence of well-recognized meanings of such words as "sugar," "chocolate," "a cupful" (standard measuring cup), "hot oven" (so many degrees, etc.), as well as of all the words designating operations like "mix," "pour," etc. The test of all this would lie in the similarity of the performance and the product when numbers of different people are provided with the recipe and follow it exactly.

Operational definitions, then, are merely definitions which consist as far as possible of words *clearly designating performable and observable operations subject to corroboration.* Thus, they may consist of (1) physical manipulations, such as baking a cake or reading a thermometer; (2) objective verbal designations of these manipulations; or (3) verbal designations of symbolical or mental operations, such as the definition of cube root or other mathematical operations.

Now the degree to which the above criteria can be satisfied varies considerably according to the stage of development which a science has reached. The prevalence of operational definitions, for example, in physics is undoubtedly due to the ready availability and standardized characteristics of the instruments for measuring time, space, temperature, voltage, etc. It is not contended, therefore, that other definitions

which are only imperfectly, slightly, or not at all operational may not be valuable in the early stages of a science, and on the frontiers of well-established sciences. They may be useful as first approximations, pending more mature developments. Highly perfected operational definitions are goals toward which to strive, rather than tools to be hoped for or to be conjured up ready-made at the outset of an inquiry. This has always been recognized by the proponents of operational definitions.

One other argument against the feasibility of operational definitions in sociology should be briefly considered. This is the familiar contention that it is not only the maturity of a science, but the nature of the subject-matter which determines whether operational definitions are practicable or desirable. In this connection, we encounter again several crucial misconceptions which we remarked upon in the first chapter. It is frequently assumed that some phenomena, especially social phenomena, are intrinsically subjective, others intrinsically objective,[11] and that consequently some phenomena are amenable to measurement and operational treatment, others not. As we have already pointed out in the first chapter, these categories (objectivity, measurability, etc.) are in science not regarded as the inherent characteristics of anything. They are merely designations of different ways in which man responds to phenomena and communicates his responses. When an observer communicates his observations so that others can corroborate his reports, we call such data objective. As for measurability, this is again merely a particular way of responding to phenomena. A large vocabulary of such words as more, less, high, low, heavy, light, etc., designate informal and non-numerical measurements.[12] The more

[11] *Cf.* John Dewey: ". . . I submit that if the words *personal* and *impersonal* were uniformly prefixed to things ('objects' in the sense in which they are linguistically identified with things) instead of the words 'subjective' and 'objective,' an artificial, because gratuitously instituted, problem would be eliminated." *The Philosophy of John Dewey,* (P. A. Schilpp, Editor, Northwestern Univ., 1939), p. 543.

[12] *Cf.* John Dewey, *Logic. The Theory of Inquiry,* (Holt, 1938), p. 202: "All comparison is of the nature of *measurement.* . . The only difficulty standing in the way of recognition of the equipollence of comparison and measurement is the fact that the results of many measurements are stated qualitatively, not in numerical terms." See also my *Foundations of Sociology,* pp. 137-42.

formal measurements are, as we shall see, *symbolic inventions* to designate gradations in different aspects of phenomena, and they are as subject to development with respect to social phenomena as with respect to any other data. As we also saw in Chapter I, this means that the imaginings, thoughts, feelings, and phenomena of "consciousness" generally, *when observed and reported in as checkable form as any other laboratory observations,* either by the observer upon himself or upon others, are as amenable to scientific study as any other body of data.

E. OPERATIONAL DEFINITIONS IN THE SOCIAL SCIENCES

We may now summarize this discussion of definitions with two concrete illustrations of the kind of work that is now being done to develop more specific, objective, and operational definitions of sociological terms. Consider such a word as "morale." What does it mean? One source gives the following definition: ". . . the degree to which the individual feels competent to cope with the future and achieve his desired goals. [Negatively it is characterized by] . . . feelings of insecurity and discouragement."[13]

Note that this definition consists of words about a word, and that the entity defined is said to be an individual's *feeling* about his situation. In other words, "morale" would appear to be one of these "subjective," "inner," "mental" phenomena frequently supposed to be beyond the reach of scientific methods. The attempt to use the above definition as it stands in an actual investigation of whether some groups have more morale than others, would doubtless result first of all in verbal quarrels over the definition. Some would suggest another wording, because as it stands, the definition "does not cover," or "leaves out" some things that "really are" morale. Since the number of different associations which different individuals may have with a word are very numerous, prolonged discussion, numerous definitions, and no conclusion as to the "right" one would be the result. Clearly, what is needed, therefore, is the designation of some behavior (ver-

[13] E. A. Rundquist and R. F. Sletto, *Scoring Instructions for the Survey of Opinions,* (Univ. of Minnesota Press, 1936). My discussion of this illustration is adapted from F. S. Chapin's excellent article, "Definition of Definitions of Concepts," *Soc. Forces,* 18:153-60, 1939.

bal, emotional, or any other) which as individual items or as a configuration, *we agree to call "morale."* These items must then be arranged in a scale or test so that people who are characterized by a high degree of the behavior we have agreed to call morale will, when measured by the test, receive a *high* score, and vice versa.

The methods of constructing such tests will be discussed in a later chapter. Suffice it here to say that the process described above was actually carried out by Rundquist and Sletto, and that with the resulting instrument they could determine objectively (*i.e.,* verifiably, in terms of the specific behavior taken into consideration by the scale) the difference in morale between, for example, employed and unemployed groups. Chapin[14] has constructed on similar principles an instrument for defining and measuring the social status of American urban families. The methods of constructing and validating this scale will also be discussed in a later chapter (Chapter IX). Our concern here is merely with the principle involved, as it bears upon the technique of defining sociological terms objectively.

It is sometimes alleged that this procedure results in a "degradation" of concepts; that the terms as operationally defined (*a*) no longer mean what they meant before, and that (*b*) an operational definition neglects some aspects of behavior which many people have hitherto included in it (*e.g.,* "morale"). Both these objections may be admitted. But they do not undermine the value of the procedure. It is undoubtedly true that words like "morale," "socio-economic status," "intelligence," etc., defined in terms of scores on a specific scale will have a very much narrower (or at least a more or less different) although a perfectly definite, meaning than these words have today in general folk usage. But this does not mean that all the folk meanings which these words now have would be denied, or ignored. We may admit that some or all of the phenomena to which these words refer in folk language and literature are possibly quite as important as those included in the objective and specific definition provided by the scale. These *other* meanings, in so far as they

[14] F. S. Chapin, *Contemporary American Institutions,* (Harper, 1935), ch. 19.

are relevant to the problem we have set ourselves, must be similarly defined operationally by *other* words, symbols, and scale scores. If the objective and operational definitions of words leave out "shades" of meaning which were formerly attached to the words defined, the solution is to develop equally clear definitions for these neglected aspects, either as a gradation of the main definition or as an independent word denoting a definite aspect.

This will mean that common words when adopted for scientific purposes (*e.g.*, horsepower, foot, water $= H_2O$, etc.) will undoubtedly have meanings quite different from their folk connotations. But this refinement — "degradation" some would call it — is the very essence of science. To argue that "the more precise and unambiguous the terms become, the less valuable they are" on account of the numerous folk connotations which scientific terms leave out, overlooks the fact that all the neglected aspects are subject to equally objective definition. The sensing and objectifying of increasingly subtle aspects of phenomena is the endless task of science.[15]

F. THE UNITS OF SCIENCE AS CONSTRUCTS OF CONVENIENCE

The same difficulties and confusion which we have observed above with respect to terminology generally, apply to the more specific problem of units of measurement in the social sciences. For, although society is a whole, with all its aspects and parts inextricably interwoven, the limitations of our senses require that we study it in parts. The same is true of every field of science. This gives rise to the difficult question, "Into what units or parts is it permissible to break up the whole?" Here again, our final criterion must be the usefulness of the units for the purposes at hand. And here again we encounter many conflicts with conventional, traditional, and metaphysical units and classifications.

One of the commonest controversies in the social sciences over the matter of units is the controversy over the "reality" of "society," or the "group," as a unit. Because of the time-honored focus of attention on the individual as the measure

[15] For elaboration of this subject, see my *Foundations of Sociology*, pp. 58-77. See also Chap. IX of the present text, pp. 302-306. See also "Is Ambiguity in Some Social Science Concepts Desirable?" *Amer. Jour. Sociol.* (forthcoming 1942).

of all things, the assumption by some sociologists of the group as an equally "real" entity, has been vigorously attacked both within the field of sociology and outside. This controversy again betrays the same failure noted above with respect to terminology generally, namely, the failure to recognize that terminology and units are merely convenient delimitations of phenomena for certain purposes of study, and that as such, a term or a unit must justify itself purely by its utility. Thus the individual organism is a bundle of parts, a functional arrangement, which we have found it convenient for many purposes to consider as a unit. For certain other purposes, however, it is wise to break up this unit into smaller parts, represented by the numerous units and classifications of physiology, psychology, chemistry, and physics. Usually these subdivisions, when first proposed, have been vigorously opposed, and their "reality" denied. It is not surprising, therefore, that the attempt to synthesize individuals into a new unit called a "society," and to claim for it the same reality as for its component units meets with vigorous opposition. The reason is the same in both cases, namely, that the new units call for a readjustment in our mental focus — our habits of thought — which is painful. It is therefore alleged that the new unit is "unreal," "artificial," and otherwise objectionable. Professor Wallis has admirably disposed of this particular controversy in the following excellent passages:

> A group is an entity when it shows persistence of qualities and when the parts of which it is composed are mutually interdependent. In such a case it functions as a unity, and so is a reality. . . The reason why the concept of a human individual is so easy, is that we start with it, not acknowledging the synthesis. In the realm of individual psychology we do the same thing.
>
> . . . I do not object to this procedure, but merely point out that it is a procedure, and that the procedure itself usually escapes our knowledge just because it is an overly familiar procedure. Far from objecting to it, I would be ready to admit that a solar system is as much a reality as is a grain of sand, and that it is as true to say that a solar system gives us grains of sand, as to say that grains of sand give us a solar system. In other words, the things discovered by analysis possess no higher reality than those discovered by synthesis. Analysis and synthesis are merely different methods of approaching the study of things, for becoming

acquainted with the nature of the world in which we live, or which lives in us.

Our justification for these conclusions will be our ability to predict social phenomena by a study of social phenomena, and the corresponding inability to predict them by a study of individual psychology.[16]

Another aspect of the same failure to recognize the purely "artificial" or utilitarian nature of *all* terminology and units is the controversy over the relationship of parts to the whole.

The question of the relationship between the individual and the group is still very much in the foreground of sociological discussion. In psychology the reverse of the same question is a matter of controversy between the Gestalt school and other factions, the "Gestaltists" being dedicated to the proposition that a study of individual responses in isolated parts does violence to "reality," inasmuch as it is the organism responding to the situation as a whole which behaves. The proponents of the so-called case method in sociology and the statisticians are haggling over the same difficulty when the case workers declaim against the statisticians' practice of lumping together large numbers and extracting an average. This average, it is solemnly affirmed, lacks "reality." There is little point or purpose to any of these discussions. Nor is there any possibility of settling the controversy as long as the terms and concepts employed are thought of as absolute entities—real in their own right—instead of convenient categories into which, for certain purposes, we have seen fit to divide the phenomena of the universe. The alleged greater reality of one category as compared with the other, is purely the result of convention and familiarity. The only justification for any unit or classification is its utility for our purpose.

That *units* of any order are merely convenient constructs, and not metaphysical entities with a mysterious reality of their own, is a fact which cannot be too strongly emphasized. For the social sciences are still largely under the influence of philosophical word-puzzles, which causes them to waste much time and energy on entirely futile discussions. Thus, in addition to the metaphysical attitude toward units noted above,

[16] Wilson D. Wallis, "The Independence of Social Psychology," *Jour. Abnormal and Soc. Psych.*, 20:147-50, 1925.

the whole-part relationship is receiving serious discussion of a type to edify any philosopher. May one analyze either the whole or its parts without destroying the reality of either? As we have seen, our knowledge of the world and its parts is conceptual, and analysis as such consists of the substitution of one set of symbols for another because it is easier in some sense and for some purposes to reason with (manipulate) one set of symbols rather than another. Thus, mathematics affords no end of illustration of the fact that, although two expressions may be "equal" or "identical," they lend themselves to widely different types of manipulation. The question whether a particular set of parts in a given relationship is to be identified with the whole depends entirely on the purpose for which the identity is to be asserted. The validity of each method of symbolization is a question of the function it performs. For some purposes, a mere algebraic sum of the parts is equivalent to the whole. For other purposes, it is quite inadequate, as further specification of the relationships of the parts may be required. To use Copeland's illustration, "as a sum of so many sticks, one cord, a neat wood pile in the back yard is identical with those same sticks scattered hit or miss over the yard. But as a whole of the type which we designate as a neat pile, there is obviously an important distinction."[17] Logical distinction must not be considered the same as physical dissection. "Conceptual analysis permits the parts to be conceived *as they are when parts of the living whole,* and not as they would be if physically isolated. . . An analysis is a veritable construction of the person who makes it. The parts are not logically in the whole. In constructing an analysis, a person builds up a new—and perhaps better—way of responding to a thing—a new way of dealing with it 'mentally' . . . The merits or demerits of a physical or any other type of analysis are for a scientist questions of the success they yield in predicting or controlling."

G. REQUIREMENTS OF DESIRABLE UNITS

With this view of the nature of units before us, we may summarize briefly some of the concrete principles of deter-

[17] Morris A. Copeland, "An Instrumental View of the Part-Whole Relation," *Jour. Phil.,* 24:96-104, 1927.

mining units for specific purposes. While it has been emphasized above that one of the principal requirements of a term or a unit is that it should mean the same to all minds, this does not mean that a generally accepted definition is the only requirement of a satisfactory unit. Four principal requirements may be summarized as follows: (1) Appropriateness, (2) Clarity, (3) Measurability, (4) Comparability. Let us consider briefly the meaning of these requirements.

1. *Appropriateness.* The appropriateness of a unit must be determined entirely by its usefulness for the purpose at hand. It follows that a unit which is appropriate for one purpose may be entirely inappropriate for another. Thus, if the purpose is to determine the increase or decrease in the amount of cultivated farm land, account of the number of "farms" would not necessarily reveal the answer sought. If it is desired merely to find out the increase or decrease in the number of farms, however, the unit "farm" is satisfactory. Likewise, if the purpose is to determine the cost of living in a particular place, a determination of the *wholesale* prices of products consumed might lead us widely astray. The appropriate unit here is obviously *retail* prices. Again, if the purpose of a study is to compare the birth rate of two cities, the population as a whole of each city is not an appropriate unit. Nor is the total female population an adequate unit. The best base for the comparison is *the number of married women of child-bearing age* (15-50 or 15-45). With respect to geographic units also, the purpose of the study determines the appropriateness of a particular unit. For some purposes, such as the computation of the density of population, a city block or a square mile may be a satisfactory unit. But if these areas contain less than 500 population they would be very inappropriate units for the computation of reliable and comparable birth, death, morbidity and other vital rates. Likewise, the city ward may be a suitable unit for certain political purposes, but it is usually very inappropriate for purposes of health and social administration.

2. *Clarity.* The importance of objective definition of units so that they will be used and understood in the same way by different people and at different times has been emphasized at some length above. The general criterion of clarity is that

a unit shall mean the same thing to all people concerned. The following illustrations of lack of clarity in units in common use will emphasize still further the need of specifying exactly the meaning of units employed:

Unit	*Possible Questions of Clarity*
Price	Retail? Wholesale? At what time? Comparability of quality of articles of same name? Value of money at different times?
Age	On date of enumeration? At last birthday? At next birthday?
Birth	All births, including still births? Live births only? Legitimate births only?
Accident	Injury causing *any* loss of time? Injury causing *specified* loss of time?
Crime	Felony only? Any offense against the law? Any offense causing arrest? Any offense resulting in conviction?
Dwelling-house	A church, if sexton lives in it? A garage, if chauffeur lives in the second story? Is a duplex one or two dwelling-houses? Hotels, apartment houses and dormitories?
Room	Does it include attic, pantry, kitchen, basement, bathroom, closets?
Residence	Where a person sleeps? Works? Eats? Location of employer?
Infant	Child less than one year? Less than two years? Less than six months? Less than one month?
Farm	"Is or is not a five-acre market garden a farm? . . . If a man owns two eighty-acre tracts half a mile apart and works them both, do they constitute one or two farms? If his hired man lives on one eighty [acre tract] does that change the status? If he rents one eighty [acre tract] to a tenant what is the effect?"[18]

How are these problems of clarity to be decided when we confront them in an actual investigation? The principal

[18] W. I. King, *Elements of Statistical Method*, (Macmillan, 1921), p. 44.

consideration must, of course, be the purposes we wish to achieve. If we are interested, for example, in the cost of living as reflected in family budgets, we decide to consider retail rather than wholesale prices. If, as is usually the case, we wish to compare our findings with those of other studies, we have to be guided largely by the definition of units used in these other studies or at least by the possibility of translating the units into terms of each other. In short, all of the questions raised above have to be settled finally on the basis of the anticipated or experimentally secured results of using the different possible definitions, in relation to the type of result we desire.

Suppose we define crime, for example, so as to include traffic violations in a congested area where nearly everyone owns an automobile. If on this basis we find the community to be highly inclined toward criminal behavior as compared with a neighboring slum where few of the inhabitants drive automobiles, the result probably does not serve the purpose for which the investigation was undertaken. If so, we should change the definition of crime so that its application in our study will serve our purpose. If a certain definition of "case" by a social agency, which measures the extent of its activities in terms of "cases," does not apply to three-fourths of the clients of the agency, this definition of "case" should doubtless be changed so as to include most or all of its clients. In any investigation, units must be adapted to the results we wish to achieve. To be sure, the more *generally* useful a unit is the better adapted it is for *scientific* purposes. But *units* are never inherent in nature or dictated by data. Units are merely convenient symbolic designations of aspects of phenomena. The assumption that such units as "man" or "sheep" are "natural" units because they lend themselves readily to a number of general purposes and to counting, is quite fallacious. The aspect of "men" and "sheep" in which we are interested (*e.g.,* the *size* of a *group* or a *herd, i.e.,* the *bulk* or *volume* of a kind of entity called "group" or "herd") when we use them as units is always implicit in the context. This implicitness in the definition of some units as compared with explicitness of definition in scientific units gives rise to an assumed fundamental difference designated by

such terms as "natural" (*e.g.*, man, sheep, etc.) *vs.* "artificial" (*e.g.*, degrees, calories, etc.).[19]

When we pass from the problem of defining with clarity such concrete terms as those illustrated above to the so-called "abstract" or "relative" terms like "normal," "crime," "pathology," "marital adjustment," etc., we are confronted with the same problem. We tend to approach these problems, however, with less confidence and in a less practical way than those discussed above because of a certain feeling of mystery and fear lest by an explicit and arbitrary definition we miss "the thing itself," "the *real nature* of the phenomenon." Yet our "feeling" as to the "nature" of the entity represented by a word is merely the sum total of the associations which we have developed for that word. We do not deny the legitimacy or the importance of these associations, including the vague and subtle ones which may be held only by ourselves. It is only by explicitly specifying these associations, however, that we can make clear to others just what meanings and associations we do attach to a word. If these meanings are very varied, we may find it in the interest of mutual understanding to agree to apply the original word only to certain well-defined aspects of the phenomenon under consideration, and to adopt other terms to cover the excluded associations. Or we may leave the original word as a non-scientific catch-all symbol for everybody's associations with a word and substitute a new word for the special aspect in which we are primarily interested.

Proceeding in this manner, we may recognize at least two meanings of the word "normal." Very frequently it is perhaps used to indicate a state of affairs which the user regards as desirable, or even ideal, as, for example, when he considers his health normal. Such a definition would obviously be relative to each individual, and would vary greatly in content from person to person. A more objective definition of "normal" is, therefore, to define it as simply the usual or ordinary for a given group as measured by a statistical mode, or the variations within, say, one standard deviation above and below the mean. Subnormality and supernormality would then represent any variation respectively below and

[19] For elaboration of this subject, see my *Foundations of Sociology*, pp. 68-70.

above this central or most common group. This assumes that a scale indicating degrees of the quality under consideration (intelligence, ability, industrial production, marital adjustment, or anything else) has been constructed and that the data have been classified according to this scale. "Pathology," "crime," "insanity," "maladjustment," and other such words are then definable as a specified degree of variation (say the lower two percent) in *the socially disapproved direction*. Corresponding degrees of variation in the socially *approved* direction are called "genius," "distinguished," etc.[20]

The fact that scales for the measurement of such phenomena are *evaluations relative to some culture* does not affect their logical foundation or utility as instruments for discriminating and categorizing the responses of men to different degrees of an aspect of a phenomenon. In this sense *all* human responses are "evaluative." That is, we "value" the weight of one article as more than that of another, and instruments exist for the more accurate discrimination and classification of these "valuations." Likewise, we "value" one liquid as sweeter, cooler, or darker than another. Thus, "valuating," which is frequently supposed to be a unique feature of our responses to social phenomena, is now seen to consist of nothing but discriminations (differential responses) which we also make toward all physical phenomena. It is true that these value scales of physical qualities usually are much more universal and interculturally valid, thanks largely to the diffusion of standardized instruments for physical measurement. But in the end, all such judgments of degrees of physical qualities, whether with instruments or without, are relative to the sensory equipment of man. Social valuations and social scales are likewise relative to this sensory equipment *as conditioned by group living* in a given culture. It is not necessary to argue that a tape measure and an attitude scale are the same *in any other respect* than as instruments through which it is possible to discriminate reliably significant gradations in situations to which we have to adjust. This process is usually called measurement.

[20] For further elaboration of this subject, see S. C. Dodd, *Dimensions of Society*, (Macmillan, 1941), Chap. 5. Also G. Lundberg, "Societal Pathology and Sociometry," *Sociometry*, 4:78-97, 1941.

3. *Measurability.* To say that measurability is a desirable characteristic of scientific units is then merely another way of designating a relatively high degree of clarity in definition. The definition of the valuative, subjective, intangible qualities of social phenomena is to be solved as the same problem has been solved in the other sciences. Height, weight, temperature, etc., are defined for scientific purposes only in terms of the instruments we have devised to measure those aspects of situations to which we need to make increasingly refined adjustments. A mature social science will likewise define intelligence, morale, status, maladjustment, crime, etc., in terms of the standardized instruments by which we measure certain aspects of situations which confront us. There will be no mystery as to exactly what segments or aspects of behavior are taken into consideration in the calibration of these instruments. Neither will there be any denial of the possible importance of other aspects which these instruments neglect, and which will therefore be legitimate objects of further construction of instruments. The technique of this construction will concern us in a later chapter.

We have pointed out that in science we take measurement to be merely a more specific and refined way of responding to events. Measurability, therefore, is regarded not as a peculiar quality or characteristic possessed by some phenomena but not by others. We regard measurability instead as a technical problem of devising the necessary instruments for categorizing our responses with that degree of reliability and objectivity which distinguishes measurement from other types of response. Definition and measurement tend in science to involve the same operations. That is, measurement is a way of defining by describing, and distinguishing accurately, just as the most accurate definition tends to be in terms of measurements. Since Einstein, at least, physicists have been content to define time as *that which* is measured by clocks, and to say that force *is* that which makes pointers move across dials, etc. Previously physicists, too, were of the impression that they must define, describe or "know" what these things "are" in some more final sense than that which their measurement yields. Following out this notion they constructed ponderous paragraphs of familiar words whose reassuring

sound seemed to provide the necessary definitions. When they faced the concrete problems of their science, however, it was found that these definitions contributed nothing to the solution. All that was found useful was *the measurements* of the entities so loquaciously defined.

Ordinarily we limit the term measurement to those descriptions which we can state in numbers of some kind. We strive in science, therefore, to state our observations in units which, when counted, represent a reliable and useful account of the event under observation.

The obstacles to measurement alleged to inhere in the nature of the subject-matter of the social sciences are merely the difficulties which inhere in the type of units into which that subject-matter is at present divided. *They are not concretely observable units,* but are assumptions and abstractions drawn from philosophy and metaphysics. The general criterion of objective measurability of a unit is the possibility of securing similarity of responses to it from most or all people. The "strength of religious interest" in an individual is not "directly" measurable. But his religious behavior (including his symbolic responses)—his church membership, attendance, financial support, and other activities engaged in, or abstained from, for religious reasons—is objectively measurable. When methods of accurately describing these types of behavior in terms of numerical degrees are devised the task of measuring them has been solved.

4. *Comparability.* Facts are of significance only in relation to other facts. It is of primary importance, therefore, that units be not only clearly defined, but that care be exercised lest the definition change over periods of time within which it is desired to make comparisons. A vast amount of error in statistical studies in the social sciences flows from a failure to observe this criterion. Comparisons of wages, wealth, and income in money units which vary widely in purchasing power between the periods and places compared are common. Numerous attempts have been made to discover trends in social maladjustment through comparative figures of divorces, arrests, convictions, etc., without taking into consideration the changes in laws, and law enforcement during the period and in the localities considered. The Fed-

eral Census itself has changed its definition of some units from decade to decade, *e.g.,* such units as family, city, farm, establishment, age, etc. Unless allowance is made for these changes in definition, conclusions drawn from a comparison of the figures are bound to be fallacious. It is undoubtedly necessary to change the definitions of units from time to time, although the resulting incomparability of apparently comparable figures is likely to result in considerable inconvenience or, more seriously, in false conclusions. If the units in each case have been defined in objective and measurable terms, however, the problem is less serious because the difference between the old and the new units can then be determined and the necessary adjustments made. For the general criterion of comparability is that the symbols representing the units to be compared should in all significant respects correspond to the aspects of phenomena which we wish to compare. Statistical devices such as percentages, ranks, index numbers, standard deviation units, and the components of factor analysis are good illustrations of instruments designed to render comparable *similar aspects* of widely diverse subject-matter.

H. CLASSIFICATION

In the above discussion of the definition of terms and units, the general principles of scientific classification have already been set forth. A definition of any kind is, of course, only a kind of classification. The fundamental reason for classification is that science is invariably concerned with discovering and describing *uniformities* in large numbers of phenomena. And the process of determining these uniformities consists of separating the whole mass of phenomena into categories on the basis of their likenesses and differences. Certain characteristics of an object are selected as significant for certain purposes. Objects possessing these characteristics are classified together; objects which do not possess the significant characteristics are excluded from the class. These likenesses and differences may be quantitative or qualitative — in terms of "place, time, circumstance, quality, magnitude, activity, behavior or function, coexistence or sequence."[21]

[21] F. H. Giddings, "Societal Variables," *Jour. Soc. Forces,* 1:346, 1923.

The number of classifications which can be made out of almost any body of phenomena on these bases is exceedingly large. What characteristics we shall use as the basis of classification is always to be determined by the end sought. "The test of the scientific character of a classification," says Westaway, "is the number and importance of the properties which can be regarded as common to all the objects included in a group." [22] But the importance of a property or an attribute will be governed by the purpose we have in view.

For example, in a housing survey, we could classify houses according to color, and secure a fairly complete classification. But both the number and importance of statements applicable to the resulting groups would be negligible. Likewise, in a study of the death rate in a population, a classification of that population as blondes and brunettes would be practically useless. The classifications which we need in this connection are age, sex, nationality, race, etc. "It often happens that natural groups must be founded not on the more obvious but on the less obvious properties of things, when these are of greater importance; and this suggests what is certainly the fact, that an extensive knowledge of the properties of objects is always necessary for making a good classification of them." [23] The difficulties of adequate classification should be fully recognized. But they are difficulties which are not insuperable, once we abandon the attempt to make classifications which will hold for all purposes, and recognize all classifications as valid only for specific purposes. "Performed consciously or unconsciously, the act of classification is indispensable to, and accompanies every scientific inference. A mind is orderly or slovenly according as it does or does not habitually and accurately classify the facts with which it comes in contact. The success of an investigation, the worth of a conclusion, are in direct proportion to the fidelity to this principle and the exhaustiveness with which the process is carried out." [24]

The formal rules for valid classifications may be summar-

[22] F. W. Westaway, *op. cit.*, p. 215.
[23] *Ibid.*, p. 216.
[24] Frank Cramer, *The Method of Darwin: A Study in Scientific Method*, McClurg, 1902, p. 98.

ized as follows: (1) Classifications must be mutually exclusive. A classification certainly loses its usefulness if any of the data to be classified can properly be placed in more than one category. This means that there must be only one basis for division at a time, and that the basis must be as objective as possible. The attempt to classify human behavior on the basis of interests, wishes, or other subjective categories, for example, is futile because a given behavior may be attributed to a number of different assumed desires or interests. (2) Closely related to the principle of exclusiveness is the principle that a valid classification must be exhaustive, that is, include all the data studied. Thus a classification of population as under twenty-one years and twenty-one and over would be exhaustive, but a classification on the basis of white or Negro might leave a large portion unclassified. (3) Finally, a classification must be appropriate, that is, suited to the end sought. Schoolboys might be classified on the basis of the number of syllables in their names. But it would probably serve no useful purpose.

The skillful classification of data is an extremely difficult and important step in the scientific method. It calls for all the imagination of the artist and the philosopher as well as intimate knowledge of the data to be classified and of the rules of logic and mathematics governing formal classification.[25] It is therefore not a subject to be adequately dealt with in any textbook, but is rather to be attained by experience and experiment with actual data. We must limit ourselves here to a simple example of the type of considerations and problems that arise in such a practical and common project as the classification of occupations. For more adequate treatment of the subject, the student is urged to consult some of the sources mentioned at the end of the chapter.

In 1938, the American Statistical Association and the Central Statistical Board appointed a committee to prepare a list of occupational titles under which the entire working population of the nation could be classified. Some of the prob-

[25] For a good example of formal classification and a discussion of its uses, see Karl Menger, "An Exact Theory of Social Groups and Relations," *Amer. Jour. Sociol.*, 43:797-98, 1938. See also, G. A. Lundberg, "Some Problems of Group Classification and Measurement," *Amer. Sociol. Rev.*, 5:351-60, 1940.

lems and questions that arose are indicated in the following extracts : [26]

The general point of view from which the Committee worked is that an occupational classification is fundamentally an instrument for dividing a working population into groups which are more or less homogeneous with respect to the occupations in which they are or have been engaged. Such groups are often used primarily as a means of *describing* a population, either for statistical purposes (as in the case of the Bureau of the Census), or for administrative purposes (as in the case of a placement office). More frequently they are used for purposes of statistical *analysis*, in which occupation is one of a number of variables. The typical analytical problem of this kind is concerned with some social or economic variable, such as employment, unemployment, income, morbidity, mortality, fertility, education, and mobility. In such problems occupation is one of a number of descriptive items normally ascertained, others being sex, color, age, residence, and place of origin. In general, the purpose of including occupation is to test or measure the relationship between the part which people play in the processes of production and distribution and their biological, social, and economic characteristics.

In order to avoid circular reasoning in such studies, it is clear that the occupational classification must be based on occupational criteria as far as possible rather than on other criteria — that it should be determined by the nature of the work done and should be quite independent of other variables. For instance, in analyzing the relation between occupation and the incidence of industrial accidents, it would weaken the argument to use an occupational classification which in itself was based in part on the incidence of accidents [*e.g.*, "hazardous" occupations]. Similarly in analyzing the relation between occupation and income it would be questionable to use an occupational classification which *directly* rested on an income criterion. . .

Deciding on the degree of detail to be obtained through its occupational classification scheme was a second important problem facing the Committee. It is an elementary fact that the final factor limiting detail is the precision of the source data, for example, whether the primary returns distinguish clearly the "welder" who is scarcely more than a machine operator from the "welder" who is a highly skilled craftsman. . .

In the opinion of the Committee, the dangers of misclassification as a result of using occupational classification schemes more refined than the data will justify, are very considerable ; and the damage cannot always be repaired subsequently by rear-

[26] P. K. Whelpton and Edward Hollander, "A Standard Occupational and Industrial Classification of Workers," *Soc. Forces*, 18:488-94, 1940.

rangement or combination. Accordingly the list of occupations prepared by the Committee is keyed to a level of precision of source data that can reasonably be expected from a household enumeration in which most of the data are obtained from persons other than the worker himself, but closely associated with him. . .

If a working population is to be classified both occupationally and industrially it is recommended by the Committee that in principle each occupation be treated as existing in several industries rather than as being confined to a single industry. Under this procedure all telephone operators are not classified under the telephone industry, where the majority work, but are distributed among the department stores, railroads, airplane factories, and other industries employing them. Exceptions to this plan are advantageous in the few cases where the inclusion of industrial terminology in an occupational title restricts that occupation to a particular industry. For example, the occupation "College presidents, professors and instructors" can be found only in the industry "Educational services," and the occupation "Servants, private family" can be found only in the industry "Domestic service. . ."

In applying an industrial classification to employees of a diversified organization an important question is the relative consideration which should be given to the activity at the place of work and to the main activity of the organization. For example, should miners in a coal mine and nurses in a hospital owned and operated by a steel company be classified industrially under the titles "Coal mining" and "Iron, steel and their products"? Careful study of the problems involved led to the conclusion that it is more significant and useful in such cases to base the decision on the activity at the immediate place of work — the coal mine and the hospital in the above cases. On the other hand, it appeared to the Committee that workers in an auxiliary establishment which has an integral technical or servicing relation to the major activity of a parent industry — as the wooden pattern shop of a machinery factory or the garage or warehouse of a department store — should be classified under the major activity.

A somewhat different problem arises in classifying employees working at establishments not operated by their employers. For example, in which industry should a demonstration girl working at a department store (retail trade industry) but on the pay roll of a cosmetic manufacturer (chemical industry) be classified? Similarly, in which industry should a musician playing in an orchestra at a hotel restaurant (hotel industry) but employed by the orchestra conductor (recreation and amusement industry) be classified? Such questions are difficult to answer, but the majority of the Committee believe that the employer-employee relationship should be the determining factor. On these bases, "industry of employer" should take precedence over "activity at

immediate place of work" in the industrial classification of employees working at establishments not operated by their employers.

In addition to such practical problems in classification, science sometimes demands classifications so comprehensive as to provide a place for all the phenomena in a given field of study. Thus, Carl von Linné (Linnaeus, 1707–78), the Swedish botanist, developed a *system* of classification for every known animal and plant. Briefly, his system consisted of placing any specimen first in a *class,* then in an *order,* then in a *genus,* and finally in a *species.*[27] This system, with modifications, is still the basis accepted by all naturalists. The symbols for designating genus and species (*e.g., Canis familiaris,* the domestic dog) are, furthermore, found in the textbooks of all nations, thus eliminating the quarrels and misunderstandings arising from difficulties of translating the exact meanings of words. The number of classifications resulting from this system is, of course, extremely large. Thus Wheeler says that there are fully 10,000 species of social insects alone, each of which may be said to have its own peculiar pattern of social behavior. This is a fact which might be considered by those social scientists who feel that the multiplicity and variety of social phenomena is especially great and hence prohibitive of systematic treatment.

I. SUMMARY AND CONCLUSION

The limitations of the human organism compel it to respond to a limited number of things at any one time. The resulting parts or units represent in no sense a "discovery" on our part of any inherent division in the whole. These parts are merely constructs of our own making—new ways of reacting to a situation. The only test of their validity is their utility in helping us to achieve whatever purpose we are seeking to accomplish. But because of the organization of our thinking from earliest childhood in terms of socially

[27] For a brief reference to Linné's work, see F. W. Westaway, *The Endless Quest: 3000 Years of Science,* (Hillman-Curl, Inc., New York, 1936), pp. 735-37. The tremendous importance of a classification like Mendeléeff's Periodic Table of Chemical Elements should also be mentioned in this connection. See *Ibid.,* pp. 442-45, for a brief discussion.

accepted units and classes, there is a tendency to overlook the fact that there is a socio-cultural, habitual aspect always inherent even in the most "natural" classes. We admit the continuity of nature in the abstract, but tend to forget this fact in our habit of looking at the universe in parts, which is the only way we can deal with it practically. As a result, the fundamental "artificiality" and indefiniteness of all units and classes are frequently overlooked. But the so-called "artificiality" and indefiniteness of all units and classifications does not make them any less important as devices through which we make our adjustments to the world.

The present chapter has attempted, first, to illustrate the necessity of a uniform terminology in any field of inquiry which men wish to discuss intelligently, or which they aspire to bring within the domain of science. Second, it was shown that such uniformity is possible in the highest degree only through the objectification of the terms and concepts employed so that they can be measured by means of mechanical devices outside of the organism, which were described in the preceding chapter. Third, some of the shortcomings of the terminology, the units, and the classifications of the social sciences with reference to uniformity and objectivity were discussed and illustrated.

A remedy for this situation was next outlined. Certain general criteria of a scientific term or concept were laid down. We emphasized the danger of assuming that the terms adopted represent any final or inherent delimitation of nature such as is implied in expressions like "the very nature of the thing," "the ultimate reality," "the thing in itself," and so forth. The morass into which such assumptions about terminology have led us in the fields of psychology and sociology was discussed at some length. One aspect of this difficulty was illustrated by the controversy in sociology over the alleged greater "reality" of the concept "individual," as compared with the concept "society." The general confusion on the subject of the relationship of the whole and its parts were shown to be the result of a general misunderstanding of the nature of conceptual analysis. Finally, the concrete requirements of satisfactory units and of classification generally were discussed and illustrated.

J. SUGGESTIONS FOR FURTHER STUDY

1. Supplementary reading for elaboration of the subjects of this chapter will be found in the author's *Foundations of Sociology,* Chaps. 2, 3, 7, and 9.

2. On operationism and operational definitions, the following literature, in the order given, will be found of interest:

a. H. Alpert, "Operational Definitions in Sociology," *Amer. Sociol. Rev.,* 3:855-61, 1938.

b. P. W. Bridgman, *The Logic of Modern Physics,* (Macmillan, 1932), Chap. 1.

c. H. Hart, "Operationism Analyzed Operationally," *Phil. of Sci.,* 7:288-313, 1940.

d. S. S. Stevens, "The Operational Definition of Psychological Concepts," *Psychol. Rev.,* 42:517-27, 1935.

e. S. S. Stevens, "The Operational Basis of Psychology," *Amer. Jour. Psych.,* 47:323-30, 1935.

f. R. B. Lindsay, "Critique of Operationalism in Physics," *Phil. of Sci.,* 4:456-70, 1937.

g. G. A. Lundberg, "Operational Definitions in the Social Sciences," *Amer. Jour. Sociol.,* Jan. 1942.

3. A brief summary of S. C. Dodd's work on the definition and classification of sociological concepts will be found in "A System of Operationally Defined Concepts for Sociology," *Amer. Sociol. Rev.,* 4:619-34, 1939.

4. On the theory and practice of formal classification, see K. Menger, "An Exact Theory of Social Groups and Relations," *Amer. Jour. Sociol.,* 43:790-98, 1938. Also G. A. Lundberg, "Some Problems of Group Classification and Measurement," *Amer. Sociol. Rev.,* 5:351-60, 1940. Regarding practical psychological aspects of classification, see P. F. Lazarsfeld, *The Technique of Marketing Research,* (McGraw-Hill, 1937), Chap. 11.

5. For practical experiments in the classification of social groups, see Dwight Sanderson, "Group Description," *Soc. Forces,* 16:309-19, 1938. Also "Preliminary Group Classification Based on Structure," *Soc. Forces,* 17:1-6, 1938.

6. For an excellent brief discussion of methods of classification in the physical and biological sciences, see V. F. Lenzen, "Procedures of Empirical Science," *International Encyclopedia of Unified Science,* Vol. 1, No. 5, pp. 31-35. Univ. of Chicago Press, 1938.

CHAPTER IV

THE PRINCIPAL METHODS OF SOCIAL RESEARCH

We do not speak of qualitative *versus* quantitative analysis. We do not seek to prove that one type should predominate over the other. Instead of dogmatizing about method at large, we are experimenting with methods in detail. In the measure of our proficiencies, we all practice both qualitative and quantitative analysis, shifting our emphasis according to the tasks we have in hand. And we are all eager to see our colleagues develop both types of analysis to the limits of efficiency in the tasks which they essay. — *Wesley C. Mitchell.*[1]

A. THE SIMILARITY OF DIFFERENT METHODS

The term science has been used in this book to designate the method of procedure which is today accepted as the method of *natural* science.[2] That is, it has been employed to refer only to those disciplines which engage in the objective classification of their data for the purpose of reducing a wide range of observed phenomena of behavior to a brief, generalized statement or formula. This is not, of course, the only useful type of systematic accumulation of knowledge. We may be interested merely in discovering and setting forth specific temporal or spatial events or sequences, as in history and geography. These disciplines are sometimes called descriptive sciences, but since all science is descriptive, such a distinction is of doubtful value. While the terms "history and geography" are usually thought of as primarily denoting subject-matter, they also imply a particular method. This method is usually thought of as differing from the method of natural science in that the method of history concerns itself chiefly with a recording of each unique datum occurring in the immediate or remote past, rather than with the quantitative classification of data for purposes of generalization. This distinction, however, it will be seen on

[1] "Quantitative Analysis in Economic Theory," *Amer. Econ. Rev.*, 15:1, 1925.

[2] The term "natural science" is here used in its commonly accepted meaning, as it applies to sciences like chemistry, physics and others. Of course all science is "natural" and social phenomena are as much so as the phenomena of chemistry or physics.

further investigation, is one of the degree to which a process is carried out rather than a difference in kind. The original observations of natural science are also observations of apparently unique events in the order of occurrence. But natural science does not stop at this point. It proceeds to ignore some of the unique aspects of its data in making classifications *according to certain similarities.* Science always *abstracts* certain specific elements or aspects of a total reality and analyzes them from the particular angle which is relevant to the immediate purpose. *From these classifications of similarities, uniformities, and sequences it makes its generalizations.* Thus, it merely *carries further* a process which in its early stages is in all important respects like the method of history.

B. THE HISTORICAL METHOD AND NATURAL SCIENCE

Even strictly historical method, however, tends to merge gradually and more or less objectively into the method of natural science. To the extent that historians go beyond the mere recording of facts in their temporal sequence, and become "comparative historians," that is, attempt to reduce their data to types, and to generalize from them, history becomes a natural science. Since there are usually present in the works of all historians such groupings of facts and generalizations more or less objectively carried out, it is clear that no arbitrary line can be drawn between natural science and other disciplines primarily historical in character. They tend to merge gradually into each other and differ chiefly in the degree to which they carry out their processes, and in the objectivity of their procedure. The same may be said of the crude common-sense method of the child or of the proverbial "man on the street." As soon as a child begins to experience likenesses and differences in his responses to his surroundings, he is engaged in observation and classification, however rudimentary the process may be in the conventional terms of scientific method. These similarity-and-difference responses will probably be in terms of number, size, shape, speed or intensity of objects, characteristics, or events. In making responses in such terms he is engaged in a rudimentary form of measurement or enumeration. From the fre-

quency of recurrences, similarities, or sequences thus experienced, every normal person generalizes more or less broadly. All that science adds to this common-sense procedure is *accuracy* and *objectivity*.

The purpose of pointing out at some length this essential similarity of different methods, is to avoid much futile controversy over the desirability of various methods at present employed in the social sciences. *To point out the greater completeness, objectivity, and accuracy of the methods of natural science is not to deny altogether the value of less complete and less accurate procedures.* Tentatively, at least, even the results of entirely random and subjective inquiry may have some value in suggesting problems and affording clues to solutions. Pure chronologies of events, even if they are in such form as to preclude the possibility of ever deducing scientific principles from them, may be of great practical value for certain purposes and essential to intelligent behavior with reference to particular situations. This is true of history as a pure record of events and of case histories of an individual, a group, or an institution. An accurate knowledge of the past of each may enable us to predict their future behavior with a high degree of accuracy. But on what is our reliance on such prediction based? Clearly, it can be only our knowledge of certain sequences in a *number of cases* which lend any validity to such predictions. (See Chap. I, Sec. J.) The degree to which these individual records become analytical, comparative and *statistical* will determine their value for accurate prediction, which is the chief standpoint from which they have any scientific value. In anthropology this transition toward the method of natural science is fully recognized. Thus Boas defines anthropology as "the science that endeavors to reconstruct the early history of mankind, and that tries, wherever possible, *to express in the form of laws ever-recurring modes of historical happenings.*"[3] (Italics ours.)

From this definition it would obviously be futile to argue

[3] Franz Boas, *Anthropology. Columbia University Lectures on Science, Philosophy, and Art,* 1907–08, p. 8. See also Clyde Kluckhohn, "On Certain Recent Applications of Association Coefficients to Ethnological Data," *Amer. Anthropol.*, 41:345-77, 1939. Also A. Lesser, "Research Procedure and Laws of Culture," *Phil. of Sci.*, 6:482-95, 1937.

whether anthropology is a natural science. It depends on the degree to which a particular worker devotes himself primarily to the "reconstruction of the early history of mankind" or to the formulation of laws of "ever-recurring modes of historical happenings." If he is primarily engaged in the latter activity he is engaged in natural science. But a rather complete accumulation of the historical facts is necessary, as is the case in every scientific discipline, before this method becomes possible. There is no question, therefore, of the relative value of the two methods of procedure. They are *different stages in the same attempt* to understand, explain, predict, and control cultural events.[4]

C. COMPLEMENTARY NATURE OF DIFFERENT METHODS

This view of different methods as different stages in a common process should serve to set at rest much futile controversy over the superiority or inferiority of different methods in the social sciences, and especially in sociology. Each has its place, and, *for a particular purpose, or at a particular stage of investigation, is best.* Instead of speaking of the historical method, the "case" method, the comparative method, the statistical method, etc., as mutually exclusive and even antagonistic, we should recognize that these methods are all legitimate stages in human adjustment-technique, the completest form of which we call natural science. The fallacy in controversies over the value of these different methods lies in regarding each as independent, sufficient, or complete in itself. Any method which achieves its purpose is valid *for that purpose.* Our methods should change as they are adapted to different ends or different stages of development of a field of inquiry. Even theological and metaphysical methods of approach, to use two of Comte's stages, have their value in certain stages of development of any field of knowledge. They may furnish working hypotheses and they furnish a tentative form of adjustment. Historical data and generalizations, of course, serve in the same way, but may also serve

[4] The works of James Harvey Robinson, James Shotwell, Charles A. Beard, Harry Elmer Barnes, and Howard Becker should be especially noted as illustrative of the contribution of the historical method to sociology. See especially, Barnes, Becker, and Becker, *Contemporary Social Theory,* (D. Appleton-Century Co., 1940), Chap. 15.

as case material for statistical classification and generalization according to the methods of natural science. In fact, from one point of view, all the raw material of natural science is historical, *i.e.,* it represents events which have taken place in the past. From the standpoint of verification, corroboration and objectivity of statement, their usefulness varies greatly according to the remoteness of their occurrence, the accuracy with which they are observed and the form in which they are recorded. Since this form in turn depends largely on the purpose in mind at the time of observation, it is often very difficult to use a body of facts compiled for purely historical purposes as data for natural science. The selective process in writing history renders much historical material of doubtful value for scientific purposes both on account of its neglect of some of the data, and the non-quantitative statement of those that are recorded. These considerations apply of course to all of the so-called life history documents — letters, diaries, biographies, etc. — as well as to more formal historical documents.

But scientific observation also consists of selecting from an innumerable list of characteristics and relationships certain arbitrarily chosen aspects of clearly defined objects for recording, and excluding all other aspects. The limitations of our mechanisms of response and symbolization (*i.e.,* our senses and language symbols) make this unavoidable. Among these limitations is the inability of the senses to grasp more than a comparatively few facts at a time. This makes it necessary for us to select for consideration only certain characteristics from the multitudinous phenomena which constitute the universe. Out of the things selected we make scientific facts, *i.e.,* generalizations. This necessity for selection gives rise to a question of fundamental importance to science, namely: How shall we determine in a particular investigation which factors to select and which to ignore?

D. THE NATURE AND FUNCTION OF WORKING HYPOTHESES

When we are confronted with a concrete everyday problem we answer this question by taking note only of those elements which we think are significantly related to our problem. What we think is significant, however, will be de-

termined by our previous experience. That is, we will proceed to look for factors which were significant in other *similar* problems with which we have been confronted. The similarity may be great or small. If our situation or problem is so strange that we are unable to compare it at all with other situations or problems to which we have been exposed, we are either unable to make any adjustment to it at all, or our adjustments will be of a blind trial-and-error type. The assumptions which we do make regarding the significant factors in a situation or problem become our *hypothesis,* on the basis of which we begin to make special note of the data which tend to bear out or contradict our hypothesis. In the everyday world, therefore, we usually construct hypotheses as to the cause and solution of all our problems on the basis of *analogy*—the assumption being that, because two things resemble each other in one or a number of points, they therefore probably resemble each other in some other point or points which are significant for our purpose.

It is because this assumption is so frequently used as a conclusion and a proof, instead of as a hypothesis to be investigated, that analogy is in such bad repute as a scientific process. Analogy is useful and defensible only when points of difference as well as of similarity are noted. Again, it is not so much the numerical preponderance of similarities or differences which are of importance, as their *significance for the purpose at hand.* When properly used, analogy is, not only in everyday experience, but in science as well, our most fruitful source of hypotheses.

A working hypothesis is necessary before the intelligent collection of data can begin. Even a general collection of data like the Federal Census has behind it as an hypothesis the assumption that the type of data which has been found useful in the past will be useful in the future. Without an hypothesis we should have to set out to record all our responses to environmental stimuli, *i.e.,* all of the perceptions of our senses. But as has already been pointed out in a previous chapter, what our senses respond to is determined largely by what they "look for." Whether we intend it or not, therefore, there will be a *selection* of data, on the basis of

our previous experience. This is all that a working hypothesis is—a clearly recognized assumption, based on previous experience, of what are probably significant factors in a situation. The only difference between gathering data without an hypothesis and gathering them with one is that in the latter case we *deliberately recognize* the limitations of our senses and attempt to reduce their fallibility by limiting our field of investigation so as to permit a greater concentration of attention on the particular aspects which past experience leads us to believe are significant for our purpose. Even if we started out by recording all of our responses, we should be compelled to make a selection from the mass so recorded before we could use these observations for scientific purposes. Such a selection would have to be on the basis of some hypothesis. The simpler and more satisfactory way, therefore, is to make the original observations on the basis of the chosen hypothesis or hypotheses.

The working hypothesis, then, is a tentative generalization based on a cursory, empirical examination of certain data. But it may be merely a "hunch" or an "intuitive" supposition. Fundamentally, of course, all "hunches," "intuitions," and hypotheses rest upon an informal, or even "subconscious," inductive basis. An hypothesis is a tentative generalization based on data so vague or inadequate that it is not at the time demonstrable in objective terms. Here, again, of course, there is much difference of opinion when an hypothesis ceases to be an hypothesis and becomes a scientific theory or a scientific law. This question turns on the adequacy of the data on which the generalization is based, which, within limits, will be a matter of opinion. Usually the term "theory" is used to designate a middle ground between the largely unsubstantiated generalization known as the hypothesis, and the fully substantiated and demonstrable generalization called a scientific law or a scientific fact. All of them are formulations which fit more or less closely the facts so far as known. The boundary lines between each are not, of course, fixed or definite. The terms represent degrees in a scale of objective demonstrability.

E. DANGERS IN THE USE OF HYPOTHESES

The working hypothesis is, in its simplest form, merely a definition of the object of an investigation. As such, it limits the problem to proportions, which can be handled with the resources at hand, and prevents an aimless and random gathering of data. It guides in the selection of data. Herein, of course, lies a grave danger. If the investigator conceives of his hypothesis as a proposition to be defended to the exclusion of evidence that might support the opposite, or another hypothesis, such an approach of course vitiates his work from the scientific standpoint. "A working hypothesis," says Chaddock, "in the scientific sense, is a theory or an explanation held after careful canvass of the known facts, in full knowledge of other explanations that have been offered, and with a mind open to a change of view if the facts disclosed by the inquiry warrant a different explanation. It is, therefore, held *with the definite purpose of including in the investigation all available and pertinent data, either to prove or to disprove the hypothesis.* An hypothesis of this character is usually desirable and even essential. It gives point to the inquiry, and, if founded on sufficient previous knowledge, guides the lines of the investigation. Without it much useless data may be collected in the hope that nothing essential will be omitted, or important data may be omitted which could have been easily included if the purposes of the inquiry had been more carefully defined. Blind gathering of masses of data does not usually lead to the discovery of unexpected relations between facts or result in new explanations." [5]

The disappointment of some students when they discover that the facts do not bear out, or contradict, an hypothesis, betrays an unscientific attitude. From the purely scientific viewpoint, a scientist's contribution is equally great whether he proves or disproves an hypothesis. Emotional attachment to a preconceived conclusion must always be avoided, for it militates against that impartial observation and collection of the data, on which depends the ultimate validity of the conclusions, as well as the usefulness of the other steps in the

[5] R. E. Chaddock, *Principles and Methods of Statistics,* (Houghton Mifflin, 1925), p. 372.

scientific method. "Hypotheses," says Westaway, "are the cradle songs which lull the unwary to sleep." Frequently we forget that an hypothesis is merely a tentative postulate, and adopt it as a datum. Much of the controversy over instincts has been a consequence of this fallacy. Frequently an hypothesis becomes a vested interest for its originator or its proponents. Their reputation, prestige, and even their employment may be threatened if new facts tending to disprove their hypothesis are admitted. Under such conditions we have the pathetic spectacle of people who started out as scientists becoming evangelists and missionaries, employing the techniques of other fanatics in their attempt to bolster up an hypothesis no longer supported by the facts.

F. CLASSIFICATION OF METHODS OF COLLECTING DATA

Methods of gathering data should be distinguished from methods of manipulating, summarizing, generalizing, or otherwise handling data already collected. The present volume is concerned chiefly with the former task. It is true that the possible ways of handling data after they are collected is largely determined by the method employed in their collection. Therefore our technique of collection must be governed to a large degree by the manner of analysis which we expect to employ. But much of the confusion and argument over current classifications of "methods" is due to a failure to distinguish between technique of *gathering* and technique of *classification, treatment, and manipulation.* Thus we hear of the historical method, the "case" method, the "statistical" method, the "survey" method, etc. There is nothing mutually exclusive in these classifications. From one point of view, all collections of data are collections of "cases." The "survey" is merely a designation of a type of investigation with reference to scope, and may utilize any or all methods of collecting the data. A useful classification of methods of gathering data, like all classifications, must specify, first, the basis or purpose of the classification; and second, its categories must be mutually exclusive.

The principal general methods of collecting data in the social sciences may be classified according to (1) scope, (2) sources, and (3) technique of collection.

With reference to the *scope* of an investigation, Professor Chapin has classified the methods of inquiry as follows:[6] (*a*) Case Work, or the study of individual cases. (*b*) Sampling, or the selection of a part of a field on the assumption that the part is representative of the whole. (*c*) Complete enumeration, or the inclusion of all the data in a field. Principles and methods applying to each of these methods of gathering data will be treated in later chapters.

With reference to *sources* from which data may be drawn there are two commonly recognized, namely:

(1) Historical Sources
  (*a*) Documents, papers, papyri, stone tablets, etc.
  (*b*) Geological strata, fossil remains, etc.

(2) Field Sources
  (*a*) Particular information from living individuals.
  (*b*) Direct observation of behavior in process (including verbal behavior when our interest is primarily in the verbal behavior, *per se*, rather than in the situations represented by the verbal behavior).

Historical sources represent the records which past events have left behind them, and which are recorded and stored by devices external to human beings. The memories of living individuals are, of course, also traces of past events, and are historical in the sense that all events which have occurred are historical. But since the purpose of the present classification of sources is to establish a basis for a discussion of techniques for collecting data, written or oral information secured for a particular purpose from living individuals is here classified as a field source. The only additional method whereby events or facts become known to us is by direct observation while they are in process. This will likewise be classified as field work.

G. DOCUMENTARY SOURCES

A careful survey of all documentary sources bearing on a contemplated inquiry should always be made before beginning a field inquiry. This is important in order to avoid duplication of work, and to suggest methods of approach—pitfalls to avoid, the difficulties involved, etc. Furthermore,

[6] F. S. Chapin, *Field Work and Social Research*, (Century, 1920), p. 47.

if we are to compare our findings with those of other investigators, a careful study of the procedure in other cases is necessary. The problem of locating relevant documentary sources in any investigation is largely one of familiarizing oneself with the various indices through which such sources may be located. The amount of documentary material available or relevant will, of course, vary greatly with the nature of the undertaking. A study of some contemporary aspect of a small, local community, such as a survey of public opinion, may require very little inquiry into documentary or historical sources, except possibly to find comparable studies in other communities containing important suggestions on methods or results. Other studies, such as, for example, a study of the trend of real wages over a period of fifty years or a study of the evolution of an institution, would require extensive library and documentary research.

No textbook can provide detailed rules for determining what data are relevant to a given problem. The person who formulates and studies a problem must judge what data are competent and relevant for its solution. Ordinarily, worthwhile interests and problems for research arise only as a result of considerable previous familiarity with a field or with a subject. If the student does not already possess such familiarity, he will probably begin by seeking expert advice on the desirable sources for a beginner. If such personal advice is not available, some readily located brief source, such as the *Encyclopedia of the Social Sciences,* may be consulted. The sources suggested in the *Encyclopedia* might next be studied. These references will perhaps in turn provide *annotated* bibliographies which will enable the student to *select* those sources which are most likely to be useful to him in the *particular* inquiry he has in mind. The amount of printed material on general subjects in any large library is usually so vast that the mere indiscriminate copying of the titles in the card catalog is an extremely arduous and wasteful procedure. Doubtless the instructor will recommend good general books to be used as a background. The general bibliographies appended to books and articles may be appended to exhibit the scholarship of the author or to suggest sources to mature scholars. They may have only limited usefulness for the be-

ginner. However, with many if not most projects it is well to get a general background of the subject first.

The student may then seek recent annotated bibliographies[7] by responsible authorities in the field, in order to save himself the dismal and frequently useless labor of examining hundreds of volumes, many quite obsolete, unscientific or barren *so far as his inquiry is concerned.* To be sure, the nature of the inquiry must determine the degree to which near and remote bibliographical material must be scrutinized. Yet, in view of the rapid accumulation of printed matter on nearly every subject, the old ideal of examining all such material is not only practically impossible, but is often pure pedantry except for the very mature scholar. For the college student *discriminating selection* of sources on the basis of what other competent scholars have found useful is essential.

A bibliography, then, should grow as a study progresses, rather than be indiscriminately and voluminously gathered at the outset from library catalogs or general reference books. The student will usually first consult the ablest authority to which he has access either in person or through such a medium as the *Encyclopedia of the Social Sciences.* He will do well next to locate through the library catalog or through periodical indices such as the *Readers' Guide to Periodical Literature* any existing *annotated bibliographies* on the subject from which he can select the sources *most likely* to be useful to him. Good recent articles may give useful leads to further sources. After perusing these sources he will be in a better position to determine how much further and in what directions inquiry is likely to be fruitful. The "Topical Summaries of Current Literature" which appears from time to time in *The American Journal of Sociology* as well as the "Abstracts of Periodical Literature," which is a regular feature of this journal, will often prove useful.[8] But these bibliographies are compiled mainly for professional sociologists

[7] Some such bibliographies are listed in the Appendix.

[8] The abstracting service of *The American Journal of Sociology* was carried on from 1895 through November 1928, when *Social Science Abstracts* took over this function until its discontinuance in 1932. The service was resumed temporarily by *The American Journal of Sociology,* in Volumes 38 and 39 (1932–33) and was regularly resumed in Volume 44 (January, 1939).

rather than for college undergraduates. For the years 1929–32 *Social Science Abstracts* constitutes for students an exhaustive review of current material in all of the social sciences except psychology, which field is covered by *Psychological Abstracts*. *The American Sociological Review* also carries from time to time selected bibliographies on various sociological subjects.[9] The annual list of "Student Dissertations in Sociology," published by *The American Journal of Sociology* and the "Census of Social Research" published annually in *The American Sociological Review* furnish additional information on recently completed or incomplete research under way in this field. Finally there should be mentioned the *Index of Research Projects* (Volumes I, II, and III, 1938–39) of the Works Progress Administration made in collaboration with the National Resources Committee and State Planning Agencies. These volumes contain a systematic classification of several thousand abstracts of recent studies conducted under the auspices of these agencies.[10] The *Catalogue of Research and Statistical Publications* of the Federal Works Agency, (WPA, March, 1941) also contains several hundred references to recent research under the Work Projects Administration.

H. CRITICAL EVALUATION OF DOCUMENTARY SOURCES

Having located the relevant documentary sources, their critical evaluation is the next problem. For convenience in evaluation, documentary sources are usually divided into two classes, namely, (1) primary and (2) secondary.

By primary sources we mean any sources providing data gathered at first hand, and compiled by the same authority under whom the data were collected from the field. Secondary sources are any other sources. Thus, the registration cards in the Registrar's office of a university would be a primary source for a study of the age and sex distribution of the student body. The compilation of these records in the university catalog would also be a primary source. A newspaper

[9] Specific references to annotated bibliographies will be found in the Appendix.

[10] Students who are not already familiar with the facilities and methods of using large libraries should read W. E. Spahr and R. J. Swenson, *Methods and Status of Scientific Research,* (Harper, 1930), Chaps. 7 and 8.

report of these figures, however, would be a secondary source. Likewise, the complete census reports of the Federal Government are primary sources, but census figures appearing in the *World Almanac* or the *Statistical Abstract of the United States* are secondary. "When the responsibility for gathering the original data and for their promulgation is undivided, whether in published form or other available forms, *the source may be called primary*. But when the authority for the data as promulgated is different from that which controlled the collection of the facts at first hand *the source containing such data may be called secondary*." [11]

The importance of this distinction lies in the probable relative reliability of the two types of sources. Primary sources are no more reliable than the persons or agencies responsible for them. But if they are really primary, and therefore not subject to verification or correction by the person using the data taken from such sources, the only criticism possible is in the nature of correction or rejection of the data as a whole because of biasing influences or faulty technique known to have attended their accumulation. Secondary sources are subject to error in proportion to the distance of their removal from the primary source. The mere processes of transcription, reprinting, etc., afford opportunities for unintentional errors. In addition there are the almost unlimited opportunities in secondary sources for misuse of the primary sources through various manipulations—partial presentation dictated by bias, rearrangements, neglect of considerations of comparability, and neglect of qualifying conditions contained in the primary sources.

Since data taken from documentary sources are subject to these numerous errors which may vitiate, largely or entirely, their value for scientific purposes, it becomes necessary to scrutinize carefully all data drawn from documentary sources. If the data are drawn from a secondary source, the first and most satisfactory method of checking them is to compare them with the primary source when that is available. When the

[11] R. E. Chaddock, *Principles and Methods of Statistics*, p. 392. Secrist points out that the distinction between primary and secondary data is largely one of degree, and also that "data which are secondary in the hands of one party may be primary in the hands of another." Horace Secrist, *An Introduction to Statistical Methods* (revised edition, 1925), p. 25.

primary source is not available, we have to apply to secondary sources the same principles of criticism that we should apply to primary sources.

The principles of criticism to be applied to all documentary sources have been well summarized by Chapin as follows: [12]

A — Documents should first be criticized as to their external or objective characteristics.

(1) Authorship should be critically examined.
(2) Sources should be critically classified.
(3) The investigator should avoid hypercriticism which makes the tool the goal and not the means to the end of knowledge of the truth.

B — Documents should then be criticized as to their internal or subjective characteristics. This is the more important division of criticism. It is analytical criticism.

(1) What did the author mean by this particular statement? What is its real meaning as distinguished from its mere literal meaning?

(2) Was the statement made in good faith?
  (*a*) Had the author interest in deceiving the reader?
  (*b*) Was the author under pressure to tell an untruth?
  (*c*) Was he influenced by sympathy or antipathy to tell an untruth?
  (*d*) Did vanity influence him?
  (*e*) Was he influenced by public opinion?
  (*f*) Is there evidence of literary or dramatic motives to distort the truth?

(3) Was the statement accurate? or more particularly:
  (*a*) Was the author a poor observer because of mental defect or abnormality?
  (*b*) Was the author badly situated in time and place to observe?
  (*c*) Was he negligent or indifferent?
  (*d*) Was the fact of such a nature that it could not be directly observed?
  (*e*) Was the author a mere witness or a trained observer?

(4) When it appears that the author was not the original observer it is necessary to determine the truth and accuracy of his sources of information.

C — Particular facts may be determined by the procedure of comparison which weighs the importance of contradictions and

[12] F. S. Chapin, *op. cit.*, pp. 37, 38.

agreements and concludes at the most with a determination of the degree of probability.

I. METHODS OF FIELD INVESTIGATION

After the general problem of an investigation has been defined; after a working hypothesis has been adopted; and after all relevant documentary sources, primary or secondary, have been scrutinized, criticized, and evaluated, the methods of conducting the field inquiry, if one is required, must be decided upon. In this preliminary work, such questions as what is the problem, what is the nature of the subject-matter, and what form of solution is sought, will have been decided. These questions will in turn determine to a large extent the *minimum scope* of the inquiry; that is, the minimum data required to justify drawing the type of conclusion sought. This is the problem (to be considered in the next chapter) of determining the smallest sample which will be statistically valid for the purpose at hand. *Other things being equal,* the larger the sample, the more reliable the conclusions. The comprehensiveness of an inquiry will be determined largely by the degree of reliability required in the results, and by the time, money, equipment, and staff at the disposal of those directing the inquiry. These considerations will also determine to a large extent the methods to be used in obtaining the data.

There are two principal ways of gathering facts from the field: (1) In the *direct method* the investigator may observe and record data by his own direct observation. (2) In the *indirect method* the investigator secures his data from *others,* perhaps from the oral or written testimony of persons having had first-hand contact with the data.

The first method is illustrated by the investigator who sets out to record the number of houses in a certain district; their size and condition with respect to heat and light; and distances to schools and churches. He may further record from direct observation the characteristics, activities, and behavior of the people living in the district. He may determine their race and nationality, their attendance at libraries, schools, and social gatherings; the extent to which they pass a certain corner, travel by certain conveyances, and visit cer-

tain stores. Finally, the investigator may by direct observation record socially important structural and functional facts about individual persons in terms of a cross-section of their present status. In this category fall the more intensive physical and psychological laboratory tests such as would be included in a medical or psychiatric examination. A number of tests have been devised to measure such things as reaction-time, coordination of impulses, emotional responses. There are also a large number of verbal tests designed to determine so-called "attitudes" and "interests."[13] Whether the latter can be classified properly as direct observation or not depends on whether the behaviors in question are regarded merely as first-hand observations of verbal behavior or whether they are taken as reliable indications of what the total behavior of the individual would be in a real-life situation. In the latter case these tests belong rather in the category of first-hand observations.

The advantages of the method of direct observation by a skilled investigator and the corresponding disadvantages of the indirect method may be briefly summarized as follows: (1) The direct method allows objective observation by a skilled observer and is subject to checking by other observers. The observations of casual witnesses are necessarily unskilled, unreliable and difficult to verify. (2) The direct method gives us an account of the actual behavior in a genuine life-situation, as contrasted with the behavior which an individual *thinks* he has performed or would perform under actual or hypothetical conditions. It eliminates errors arising from misunderstandings and misinterpretations of words and the bias likely to be present in a verbal report secured from a person supposed to have had access to the original facts, but who must rely on his memory for many of them. Our knowledge of the limitations of associative memory and the rôle of suggestion and fantasy in human behavior must cause us to accept all such reports with many reservations. Finally, (3) the direct method is the method of physical science, and has

[13] For a summary discussion of these tests, see Chap. VIII below. See also, Kimball Young, "The Measurement of Personal and Social Traits," *Publ. Amer. Sociol. Soc.*, 21:93-100, 1927. Also G. W. Allport, *Personality*, (Holt, 1937), Chap. 16. For a comprehensive list of such tests, see W. A. McCall, *Measurement*, (Macmillan, 1939), Chap. 7.

all the advantages of accurate observation and measurement. Whenever and wherever the nature of the investigation and the resources of the investigator permit, the direct method should be employed.

The method of direct observation, however, has certain limitations which have compelled most social investigators up to the present to collect their facts by the second method mentioned above, namely, the indirect method of inquiry from human beings. This is largely due to the difficulties of controlled experimentation in the social sciences, referred to in a previous chapter. Through his controlled experimentation in the laboratory, the physical scientist can arrange to have the behavior in which he is interested occur repeatedly on a small scale under conditions which assure that his observations will be valid for the phenomena under natural conditions. He can, furthermore, make observations without having to wait for the fortuitous occurrence of the phenomena in nature. He can even shorten the time space during which a process takes place, by providing artificial environmental conditions. Thus, artificial weathering conditions can be arranged in the laboratory which will test in one year the weather-resistance of a paint as fully as ten or more years of exposure to natural weather.

The social scientist, on the other hand, lacking the control of the laboratory in the observation of his data, must to a large extent await the fortuitous occurrence of the data he is seeking, or rely on reports from the memory of other people, or on historical records. Most of the facts gathered by the Federal Census, as well as by other governmental agencies, are secured in this second-hand way. A person's age, race, and nationality and religious affiliation, as well as the statement of his possessions, income, and business dealings, are secured largely from the person interviewed. There is little attempt at checking of his statements because some of them are not subject to check, and the time, labor, and expense involved in checking others would be prohibitive. We take the individual's word for the truth of his statements, although we know that, leaving out all questions of intentional falsification on the part of the reporter, his memory and knowledge of his own affairs will enable him to give only

approximately correct answers. At the present time, therefore, we rely on the untrained reporting of the great mass of individuals of society for most of our social data. Until we develop a very much more comprehensive and efficient system of official recording of significant events as they occur, we shall have to continue to rely on this personal testimony. As was pointed out in the second chapter, however, experimental laboratory techniques with the aid of motion pictures may become an important source of certain kinds of sociological data.

The fact, then, is that at the present time and for an indefinite future, we must continue to rely on others, on individual testimony for a great many of the data of sociology. This makes it of the greatest importance that we develop techniques of securing from people at large reliable second-hand information which will give a high measure of predictability of behavior without the elaborate, time-consuming, and costly process of first-hand observation of the behavior at the time it occurred. Toward this end some methods are very much more effective than others. It is possible through a careful checking of items of information against each other and against objectively established facts to determine with a high degree of reliability the accuracy of an individual oral or written report. Since it is unwise to confine our efforts to observing directly only the concrete, first-hand, non-verbal behavior of people, our problem becomes one of developing a technique for eliciting reliable information from others, from the persons in first-hand contact with the facts. This technique will usually take one of four forms: (1) Life history documents—autobiographies, diaries, confessions, letters, etc. (2) An informal oral interview (individual or group). (3) A more formal filling out by the investigator of a schedule on the basis of answers to questions. (4) A questionnaire filled out by the informant without personal supervision of the investigator. These methods are treated in Chapters VI, VII, and VIII.

J. SUMMARY AND CONCLUSION

The general viewpoint of this chapter has been (1) that all questions of method are relative to specific ends, and (2)

that the broader the purpose, the more composite must be the method adopted. The purpose of natural science being to reduce the phenomena of behavior to brief laws or formulas which will be generally valid within their field, the method of natural science may be regarded as the broadest in scope. This method, therefore, calls for the fullest combination of all techniques which produce data relevant to the complete generalized description of the phenomena under consideration. Thus it calls for the use of hypotheses. In their earliest and most tentative form these may be nothing but "intuitions," "hunches," "guesses," or other vague forms of "subconscious" induction. It may next call for a careful inquiry into the historico-genetic aspects of the phenomena under consideration, especially when these phenomena are human and social. This part of the method of natural science is usually referred to as the historical method. Then we are interested in an objective description of the present facts — a cross-section of each type of phenomenon as it is and behaves at the present time. We are interested in patterns of behavior, the presence or absence of factors, the degree or amount of each factor present, etc., in a given time, place, or situation. This is frequently called the case method. But each or all of these methods do not reduce the data to the category of science until conclusions of general validity can be drawn from the data so that reliable predictions of behavior can be made. Whether this generalization be formal and mathematical or informal and subjective, it will be quantitative in nature, and is the *sine qua non* of scientific method.[14] It does not of itself constitute a complete scientific method, however, as it rests upon the previous steps in the process. In view of this fact, it is obviously futile from the standpoint of science to discuss the relative merits of different stages in the process. What may be discussed with profit is the degree of objectivity and accuracy of different methods of performing each step. We turn now to a consideration of some aspects of this problem.

[14] For elaboration of this position, see G. A. Lundberg, *Foundations of Sociology*, Chap. II. See also Chap. I, Sec. I, of the present text.

K. SUGGESTIONS FOR FURTHER STUDY

1. A profound treatise on the fundamental nature of all inquiry with special reference to scientific method will be found in John Dewey, *Logic. The Theory of Inquiry,* (Holt, 1938). Parts I and IV are the least difficult. Chapter IV on "Common Sense and Scientific Inquiry," is especially relevant in the present connection.

2. A digest of examples of different methods that have been employed in actual investigations in the social sciences may be found in S. A. Rice (Editor), *Methods in Social Science* (University of Chicago Press, 1931).

3. Suggestions on systematic note-taking and the use of libraries as well as an extensive list of sources of information will be found in W. E. Spahr and R. J. Swenson, *Methods and Status of Scientific Research,* (Harper, 1930), Chapters 7 and 8.

4. For a lucid discussion of "The Art of Note-Taking" and for ideas on securing information from "The Written Word" and "The Spoken Word," see Sidney and Beatrice Webb, *Methods of Social Study,* (Longmans, 1932), Chapters IV, V, and VI.

5. For a fuller discussion of the historico-genetic method mentioned in the text, see Barnes, Becker, and Becker, *Contemporary Social Theory,* (Appleton-Century, 1940), ch. 15. In this connection the student should also read ch. 2 of the same volume on "Constructive Typology in the Social Sciences," the main thought of which is entirely compatible with the views of the present volume. (See G. A. Lundberg, "Some Problems of Group Classification and Measurement," *Amer. Sociol. Rev.,* 5:351-54, 1940). See also, P. F. Lazarsfeld, "Some Remarks on the Typological Procedures in Social Research," *Zeitschrift für Socialforschung,* 6:119-49, 1937.

CHAPTER V

THE SAMPLE IN SOCIAL RESEARCH

In actual practice, in science and in practical life we rely on what may be called the Principle of Fair Samples, that is to say, the belief that, with reasonable care, it is possible to judge the character of a large group, or of a whole class of phenomena, by the aid of a sample or a selection from it. This principle is sometimes called the law of Statistical Regularity. — *A. Wolf.*[1]

A. THE GENERAL USE OF SAMPLING

We saw in the preceding chapter that all observation is necessarily selective. It was pointed out that we select for consideration only certain phenomena from the innumerable characteristics and relationships that are actually present. Out of the observations so selected we make whatever generalizations the data justify. But the data on which we base most generalizations in practical life are only a part — *a sample* — of all the relevant data. The same is true of all the generalizations of science. These generalizations, both in science and in practical life, rest upon the assumption that what has been found characteristic of the sample data actually observed will also be characteristic of the whole body of data which we cannot observe. To the extent that this assumption is valid, a generalization based on a sample will be as valid as a generalization based on all the data. To the extent that this assumption is false, our generalization will be invalid.

Except for the economy of time and effort which the sampling method affords, the advance of science would be insuperably handicapped. To study all the data in any field would be prohibitive. But, through the sampling method, however, we are able to make highly accurate generalizations about the whole from a part. If properly selected, according to rules to be discussed later, ten men might, for example, give us as true an average of the height of ten thousand men, as an actual study of the ten thousand. Likewise, under certain conditions, and for certain purposes, a gallon of

[1] *Essentials of Scientific Method,* (Macmillan, 1925), pp. 114-115.

water might be truly representative of an entire reservoir. But only if certain rules and cautions are observed in the selection of these samples do we have assurance of their representative nature. The principles governing the selection of such samples are our chief concern in the present chapter.

B. CONDITIONS GOVERNING THE VALIDITY OF A SAMPLE

The degree to which the assumption that part of the data is representative of the whole is valid will be determined by two considerations, namely, (1) the nature of the data observed, and (2) the method employed in selecting them.

If the data are highly homogeneous, that is, if the differences between the various items composing the whole body of data are negligible, then any item or group of items is representative of the whole. Hence, the problem of sampling in a field that is known to be highly homogeneous is very simple. The high degree of homogeneity existing in some classes of data may result from several different causes. The process of production may so thoroughly mix the component elements of a universe that all elements will be found present in every part in practically the same proportion. For example, a teaspoonful of many foods and liquids may be a representative sample of a barrelful of the same article provided the matter has been thoroughly mixed, and provided the component elements are susceptible to such mixing. Likewise, one square inch of a given cloth is usually considered a representative sample of all the cloth in a particular bolt, though the latter may contain hundreds of yards. The process of weaving and dyeing in this case assures us of a high degree of homogeneity of different parts of the cloth. We make the same assumption when we accept a few bars of soap, a few cookies, or a few packages of a product as representative of a whole consignment. The assurance is that the entire quantity contains the same ingredients in the same proportion. When a cook tastes food while preparing it, he is working on this assumption. When the essential nature of the data insures a degree of homogeneity so high that any differences that may exist between different parts or items are negligible for our purpose, the problem of sampling is simple. A part is closely representative of the whole.

In the field of social research, however, we usually wish to study the aggregations of phenomena lacking the homogeneity we find in such cases. Stephan has summarized the problem, and some methods of dealing with it, as follows: "Heterogeneity may be defined roughly as the presence of subdivisions or groupings of cases such that the cases vary less within one or more of the groups than they do in all groups taken together, the difference in variability being greater than one might attribute to chance. The processes of economic, social, and racial segregation and other ecological processess which characterize American cities produce heterogeneity, *i.e.*, a grouping of people and various social phenomena into clusters of similar and supplementary units. This clustering by race, religion, nationality, economic status, occupation, and by many attitudes and preferences, is such that no cluster is quite like the entire community, and consequently no one cluster can represent it accurately. Information must be secured from each cluster to provide a faithful picture of the whole. Usually there are areas of admixture, but even these areas are not representative. Often there are no clearly defined groupings because the heterogeneity is a matter of continuous change along a gradient shading off from a predominance of one type of unit to a predominance of another, or from one extreme variant type to its opposite. This lack of clear cut divisions complicates the selection of a sample which will be representative of all the varieties present in the community. It is like drawing balls from a large bin into which the contents of a number of different urns have been emptied without being stirred.

"If all social phenomena were distributed at random geographically, it would be quite satisfactory to take as a sample those cases which are near at hand. However, the presence of clustering makes it necessary to take some cases from each cluster, no matter how remote or inaccessible. This introduces serious practical problems of locating the clusters, transporting investigators to each district, supervising a scattered field force, and securing information from certain types of informants. In most cities, and in rural areas also, small groups of families whose behavior and characteristics differ widely from other groups are located at relatively in-

accessible places. Another kind of inaccessibility, typified by the dog in the yard, the locked door, pretended inability to speak English, and feigned ignorance, is often associated with geographic inaccessibility, although it is also found in very accessible places. Still a third kind of inaccessibility is presented, for example, by a childless couple who live in an apartment where no one knows them, working during the day, eating their meals out, and spending only a few hours at home to sleep." [2]

Our chief practical concern with the problems of sampling in the social sciences, therefore, is a consideration of the various methods and safeguards by means of which the highest possible degree of validity in the sample may be attained. Only some general principles and a few illustrations can be given in a textbook. The application of these principles to concrete new problems must be left to the ingenuity and resourcefulness of the individual researcher who must adapt his methods to suit his material, his purpose, his situation, and his resources. Common sense and ordinary logic rather than a blind reliance on rules and formulas are essential. A wide familiarity with the practical experience of other researchers is likely to be most useful. Accordingly, the student is advised to study the cases mentioned at the end of this chapter, and others, for insight into the practical problems of sampling.

C. METHODS OF SAMPLING

The requirement of a representative sample is that a part of the whole should have, in their relative proportions, all the significant characteristics of the larger group. A number of devices are employed to achieve this result. All of them rest on the principle of *random* selection. The term *random* is frequently considered as meaning careless, unpremeditated, casual, haphazard, and "having no definite aim or deliberate purpose." [3] This is an unfortunate interpretation of the term, and, if adopted, may give unscientific results. For we aim in a random selection so to arrange conditions

[2] F. F. Stephan, "Problems of Sampling Procedure," *Amer. Sociol. Rev.*, 1:572, 1936.
[3] Funk and Wagnalls, *New Standard Dictionary*.

that every item in the whole body of data from which we wish to select the sample shall have the same chance of being selected as any other item. This usually calls for careful planning and orderly procedure, especially in exposing the whole field to the selection. It means insuring the equal operation of the method of "chance." For example, selecting names from a directory by opening it at "random" and taking the first name upon which the eye falls, might appear to be a truly random selection. But if the directory has been in common use, it is likely to open more readily at some pages than at others. The result might be a bias in favor of the more popular persons or businesses. Likewise the eye might be more readily attracted by long or peculiar names. Both conditions violate the requirements of a sample truly "random," namely, that all possible items shall have an equal chance of being selected.

The general types of sampling may be classified as follows: [4]

1. Random selection, under which each item in the universe is *assured an equal chance* of being selected.

2. "Purposive" (proportional, stratified) selection, under which selection is governed by known characteristics of the whole universe so as to secure a similar distribution in the sample. (See Table I as an example.) Thus a small local sample of families could be selected so as to correspond completely to the different sizes of families in the country as a whole, as shown by census figures.

3. Mixed sampling, which combines (1) and (2).

1. *Random selection.* — (*a*) Random selection is "often sufficiently secured by the process of spreading out a consignment of goods, etc., and marking one here and another there, avoiding the first, the last, and the most obvious, and testing the objects marked." [5] If the objects drawn in this way are apparently similar, *e.g.,* cards or capsules containing a number, the method is satisfactory. If, however, the objects are a consignment of goods or other objects toward which subtle and perhaps entirely unconscious preferences or preju-

[4] G. U. Yuale and H. G. Kendall, *An Introduction to the Theory of Statistics,* (Griffin, 1937), p. 336 ff.

[5] A. L. Bowley, *An Elementary Manual of Statistics,* (Third Edition, McDonald and Evans, London, 1923), p. 57.

dices may exist in the person making the selection, the procedure may result in a biased or non-representative sample.

(*b*) To avoid the possibility of such bias, it is preferable to secure equal chances of selection by "numbering the whole groups consecutively, writing down the numbers on tickets and shuffling them, and finally drawing at random some of the tickets and examining the objects with corresponding numbers. To avoid the writing and drawing, digits are sometimes selected at random from mathematical tables and used as if they were numbers drawn at random." [6] Drawing numbers in capsules that have been thoroughly mixed in a rotating urn is essentially the method employed in all lotteries, and is employed in selecting men for the draft army.

(*c*) Another method is to select, for example, every tenth case, one every two inches, etc., from a series, or from an alphabetical list or file. Any such arbitrary selection will usually give a representative sample provided the basis of the arrangement is unbiased with reference to the characteristic for which the sampling is being conducted. If such sampling is undertaken from a file, care must be exercised lest some records be temporarily missing. Active cases may differ significantly from inactive cases. Again, if some cases consist of more than one folder or card, they will have more than an equal chance of appearing in the sample.[7] Bias may result also from certain characteristics of the material itself. For example, sometimes all the names beginning with a certain letter are taken as a sample. If we are sampling for nationality, this method would be biased, as a disproportionately large proportion of names in some nation-

[6] *Ibid.* For a fuller description of this process, see the account of the use of Tippett's numbers in G. U. Yale and M. G. Kendall, *op. cit.*, pp. 341-44.

[7] For further details on possible errors of this kind and ways of avoiding them, see F. F. Stephan, "Practical Problems of Sampling Procedure," *Amer. Sociol. Rev.*, 1:570, 1936. See also the same author's account of the problems of sampling involved in the selection, in the Census of 1940, of a sample of two persons out of every forty (five percent) by designating certain lines of the census schedule as sample lines. For example: "If every sheet started with a new family, all the individuals on line 1 would be heads. When this is the case a sample that includes lines 1, 21, 41, and 61 will greatly overstate the number of heads, and a sample that consistently excludes line 1 will understate the number of heads. Similarly, one that consistently includes line 2 will overstate the number of married women." "The Sampling Procedure of the 1940 Population Census," by F. F. Stephan, W. E. Deming, and M. H. Hansen, *Jour. Amer. Statis. Asso.*, 35:623, 1940.

alities begin with a particular letter. Thus Stephan found that in Pittsburgh names beginning with F and T formed a seven percent sample quite representative of the entire population in economic status, race, and nativity. He warns that these results would not hold for a community with a different nationality composition.[8]

A slight variation of the above method is the selection of the items representing the maximum, the minimum, the median, the quartiles, and two deciles of the sample represented as an array. In this way a good estimate could quickly be obtained, let us say, of the grade distribution of a large class. This might form a reliable basis for comparing it with another class.[9]

2. *"Purposive" selection.*—The above methods assume that the composition of the total group is not known, and that therefore a strictly random selection, or a selection by regular intervals, will come closest to giving us a representative sample. Yet frequently something is already known of the universe from which we wish to select a sample. If this knowledge is of a general kind, "investigators may prefer to choose what they believe to be typical groups, and examine them in detail. Thus investigations as to the wages, etc., of agricultural laborers have been conducted by selecting some forty districts throughout the country, so as to include types of all kinds of agriculture, and of all economic situations. This method results in an accurate and intelligible picture, but there is no easy means of calculating any average, or of knowing the distribution by number of persons earning various rates of wages. For filling in details where the general results are known the method is to be recommended."[10] A sample of this kind is likely to approach the average of the universe, whereas the random sample may vary widely from the average. On the other hand, the random sample becomes more and more representative (*i.e.,* it gives a truer

[8] F. F. Stephan, "Practical Problems of Sampling Procedure," *Amer. Sociol. Rev.,* 1:571, 1936. This author also makes some interesting suggestions for research in this field which would permit the use of addressograph lists in setting up samples for different purposes. Stephan also suggests the possibility of combining lists with known biases in such a way as to cancel the biases.

[9] A. L. Bowley, *op. cit.,* p. 62.

[10] *Ibid.,* p. 63.

picture of variations in the whole universe) as it becomes larger, whereas this need not be true of the "purposive" sample, because it has been deliberately chosen to approximate the average. This more or less *a priori* method of selecting what is typical, is, of course, subject to the bias of the investigator. But when the typicality of the groups selected can be demonstrated, this is an excellent method.

In some cases the more or less exact composition of the total group with respect to some significant characteristics is known before we select our sample. In such cases, we can increase the chances of selecting a representative sample by selecting sub-samples proportionate in size to the significant characteristics in the total. By thus selecting a sample that is mathematically absolutely representative with regard to some significant characteristics we increase the possibility that it will also be characteristic in the other respects in which we are interested. This sampling technique, sometimes called proportional or stratified selection, is especially important in the social sciences and can best be described by an illustration.

Suppose that in a certain university we know that the total registration of 5375 students in three colleges is distributed as shown in Table I. Assume that we wish to select the smallest possible representative sample of this group of students. The smallest single group in any of the subdivisions for which we have information is the women in the senior class of the school of business administration with a total of 50. This group must have at least one representative in the sample. Practical purposes will, in fact, as we shall see later, usually demand more than a single representative of this smallest sub-group. If we assign one representative to this group, however, we should have to assign one and one-half representatives to the junior women of this school (who total 75), to be strictly proportional in our selection. Since this cannot be done on account of the nature of the units (human beings) we must increase the representation for each group sufficiently to give us whole numbers. That is, we must assign to the smallest group a number of representatives sufficiently large to give each other group representation according to its size in terms of whole units. To accomplish this,

TABLE I

Composition and Proportioned Sample of a Student Body

College	Total		Men		Women	
	No. of Students	*Min. Sample*	*No. of Students*	*Min. Sample*	*No. of Students*	*Min. Sample*
Total	5375	215	2825	113	2550	102
Liberal Arts	*2300*	*92*	*1500*	*60*	*800*	*32*
Freshmen	875	35	600	24	275	11
Sophomores	600	24	400	16	200	8
Juniors	475	19	300	12	175	7
Seniors	350	14	200	8	150	6
Business Adm.	*1075*	*43*	*700*	*28*	*375*	*15*
Freshmen	450	18	300	12	150	6
Sophomores	300	12	200	8	100	4
Juniors	200	8	125	5	75	3
Seniors	125	5	75	3	50	2
Education	*2000*	*80*	*625*	*25*	*1375*	*55*
Freshmen	600	24	200	8	400	16
Sophomores	525	21	175	7	350	14
Juniors	475	19	150	6	325	13
Seniors	400	16	100	4	300	12

we must find the greatest common divisor of all the numbers in our table. In the present case, we find that this common divisor is 25. We accordingly give each group representation in our sample in a ratio of 1 to 25. That is, we take a four percent sample. This gives each group the representation indicated in the table. The total sample of 215 would then be strictly representative of the student body in these three colleges with regard to sex, college, and class. This method of selection could, of course, be indefinitely extended to take cognizance of any known groupings in the student body.

The final selection of the members to comprise the sample would, of course, be made according to one or more of the methods of random sampling described above. The concrete procedure would be somewhat as follows:

(*a*) Determination of the significant known groupings with the number of persons in each.

(*b*) Determination of the minimum size of the sample. The smallest number which will include all the elements

represented in our known groupings would, of course, be one representative of our smallest group, and all others in proportion. But if this number makes a proportionate representation impossible, we have to select the highest ratio which will achieve this result. This ratio can be determined by finding the greatest common divisor of all the numbers representing our lowest sub-groups. In our illustration this number was 25, and each group would therefore be represented in the sample in the ratio of 1 to 25. This procedure is, of course, a purely mechanical one for determining the *relative* size of the samples from different groups. The *absolute* minimum size of the sample would have to be determined by practical considerations to be discussed below.

(*c*) When the number to be selected from each group has thus been determined, this number of items would be drawn according to any of the methods of random sampling described in an earlier section. Thus in our illustration, names of all the female freshmen in the college of education (400) would be placed on cards and shuffled. Sixteen names would then be drawn at random or at regular intervals. The same process would be followed for each group. The combination of all of these drawings or selections of sub-samples would then be a representative sample of the whole body of data considered. This sample would be strictly representative with respect to the elements taken into consideration in the selection — in our illustration: sex, college, and class.

By inference we might also consider it representative in many other ways. Thus we might expect such a sample to be typical for nationality, income, and religious affiliation of the student body. But in the absence of previous knowledge of the range and distribution of these characteristics in the student body, we could test the proposition only by one or more of the methods to be discussed in the next section. In the illustration used, the principle of *relative* size of sub-samples has been observed. The *absolute* size of the sample which we might actually select for a given purpose, however, might necessarily be larger than the minimum at which we arrived above. As a principle of absolute size, Chapin has stated that each sub-class "should be large enough to represent the character and range of variation among all the items

of this particular class."[11] Thus in our illustration we included only two items from our smallest group (senior women in the college of business administration). If we expect to use our sample as representative of this group in nationality, income, and religious affiliation, a sufficiently large number should be included at least to cover the *range* of nationalities, religions, and incomes known to exist in the student body. This would probably necessitate including eight or ten persons from this group, and a correspondingly large sample from the other groups in proportion to their respective sizes. "Except when a census is by definition merely an inventory, small numbers in the cells of a table, even when derived from a complete tabulation, have no more significance than they would if they were derived from a sample."[12] So the absolute size of the sample must be determined by a variety of considerations, chiefly with reference to its purpose and to our knowledge of the composition of our whole group. "When the sample is to provide estimates for small or highly segregated items, such as the higher income classes, for example, it is often desirable to augment the sampling for this group by selecting several times as many cases as would be selected otherwise. Theoretically, the sample should include an equal number of cases for each important subgroup which is to be analyzed separately unless there is considerable difference in variability among subgroups. If the variability of a subgroup is greater than that of other subgroups, it requires a greater number of cases in the total sample so that estimates for all subgroups will be equally reliable. This is not always feasible in practice, but the sample which includes all subgroups proportionally is not the best sample (apart from practical considerations) unless the subgroups are of the same order of variability and are either of comparable size or *not intended to be used separately in the later stages of the study.* When subgroups are selected with different sampling ratios, appropriate adjustments or weightings are applied to each in making up estimates for the entire aggregate."[13]

[11] F. S. Chapin, *Field Work and Social Research,* (Century, 1920), p. 122.
[12] F. F. Stephan, W. E. Deming and M. H. Hansen, *op. cit.,* pp. 629-30.
[13] F. F. Stephan, "Practical Problems of Sampling Procedure," *op. cit.,* p. 576.

3. *Mixed sampling.*—When the items of the sample are selected *at random* from individual groups which already have been chosen "purposively," proportionally, and/or from stratifications of the universe, this has been called the mixed type of sampling because it employs both the random and the purposive methods. Thus, if the representatives of each of the smallest subgroups in the above illustration of a student sample are selected by strictly random methods, that illustration would serve also as an illustration of the mixed method. Another illustration can be drawn from the current polls of public opinion. The more reliable of these polls have abandoned the practice of sending post card inquiries to large numbers of people selected from telephone directories or other sources, which was mistakenly hoped would give a "random" sample. Instead, the population of the country is first "stratified" according to geographic location, sex, rural, urban, (further divided according to size of city), age, income, and party affiliation. The sample is then selected as nearly as possible at random within each of these groups, and in their proper proportion, so as to correspond to the whole voting population in all of the above respects. That is, if one-tenth of all the eligible voters in the country are Republican males, aged 30 to 45, living in New England cities of 25,000 to 100,000 population, with incomes between $3000 and $5000, one-tenth of the people in the sample chosen should have these characteristics. By this process, results accurate within 3 percent[14] on most subjects can be secured from a sample of 5000 (about .01 percent of the voting population), whereas a sample of over two million by the post card method as employed by the *Literary Digest* in 1936 erred by more than 10 percent in its estimate of the popular vote.

D. THE SIZE AND RELIABILITY OF THE SAMPLE

We have been primarily concerned in the present chapter with the *representativeness* of a sample rather than with its adequacy. Adequacy deals with the variation we may expect owing to the chance conditions of sampling. As such it

[14] Elmo Roper, "Sampling Public Opinion," *Jour. Amer. Statis. Asso.*, 35:331, 1940.

is a problem of mathematics, whereas representativeness is more of a problem in applied logic.[15] The problem of adequacy is very directly connected with the size of the sample and the theory of probability. For a full discussion of these subjects, the reader is referred to any standard text on statistics or mathematics.[16] We treat here only its more elementary aspects.

Any random sample, no matter how carefully selected, tends to differ to some degree in its composition from the composition of the whole body of data from which it is selected. This is due to the operation of chance in selection. The *size* of this chance variation depends on several things. First, the greater the homogeneity of the data, the less will be the tendency of any sample, regardless of its size, to vary from the whole. This is necessarily so, since homogeneity means that the range of variability is small. Obviously a sample selected from a given age-group of males would show less variability than successive samples of like character and size selected from the whole population. That is, the *probable error* of the mean in the latter case would be greater. Second, the degree to which the average of a sample tends to vary from the average of the whole body of data depends, other things being equal, on the size of the sample. If we wish to determine the average height of 100 men, our chance of closely approximating the true average is considerably greater if we select a sample of 45 than if we select 5. In fact, it can be mathematically demonstrated that in this illustration the larger sample is likely to be three times as accurate as the smaller.[17] Assume, for example, that the difference between the mean of the five cases and the mean of the 100 cases was 3 inches. To reduce this difference to 1 inch, it would be necessary to take not 3×5 cases, but 9×5 or 45 cases. (The reliability of the sample increases directly as the square root of the size of the sample.) That is, the probable error (the deviation from the true average of the 100 men)

[15] *Cf.* R. E. Chaddock, *op. cit.*, pp. 213, 384.
[16] *Ibid.*, Chap. 9.
[17] By the formula for the Probable Error of the mean $\left(0.6745 \frac{S.D.}{\sqrt{n}}\right)$ See R. E. Chaddock, *op. cit.*, ch. 11, especially pp. 237-46, or any standard text on statistics.

of the average calculated from the 5 men would be likely to be about three times as large as the probable error of the average calculated from the larger sample.

In somewhat the same way the standard error of the proportion shown by a random sample (from a definite universe) may be computed by the formula[18]

$$\text{Standard Error} = \sqrt{\frac{p'(1-p')}{n}}$$

in which

n = the number of items in a sample from a universe.

p′ = the proportion in the sample having the given attribute.

By this formula it can be demonstrated that unless the proportion of items in the universe having some specified attribute is very small, the *chances are two out of three* that the actual proportion in the whole group, the universe, will not differ from p′ by more than the amount given by the formula. For example, if in a sample of sixteen students, eight, or 50 percent, are found to be freshmen, then the chances are two out of three, assuming random selection of the sample, that the proportion of freshmen in the entire student body does not vary from 50 percent by more than 12.5 percent. The 12.5 is obtained by inserting the value of p′ (½) and the value of n (16) in the above formula, thus:

$$\text{S. E.} = \sqrt{\frac{\frac{1}{2}(1-\frac{1}{2})}{16}} = \sqrt{\frac{\frac{1}{4}}{16} \text{ or } \frac{1}{64}} = \frac{1}{8} \text{ or } 12.5\%.$$

Hence the chances are two out of three that the true proportion of freshmen lies between 37.5 and 62.5 percent of the total.

The theory of probability, the derivation of probability formulas, and their application to concrete cases are beyond the scope of the present discussion, and must be left to statistical and mathematical treatises on the subject. We are

[18] See Yule and Kendall, *op. cit.*, Chap. 19. The "Probable Error" is given by the same formula as in the text above, if the term 2/3 is inserted before the radical. The interpretation would then be that the chances are even ("50-50") that the proportion of freshmen in the entire student body does not vary from 50 percent by more than 8.5 percent.

interested in the present connection in pointing out merely the nature of the relationship between the size of the sample and the degree of stability in the measures obtained from the whole universe or body of data. *Neither the actual size of the sample nor the size of its probable error necessarily guarantees its representativeness.* As Chaddock has shown, all the data in the universe from which the sample is drawn may have a bias which a mere increase in the size of the sample cannot, of course, eliminate.[19] This bias can be eliminated only by considerations of logic in the technique of defining the universe and selecting the sample. In the same way, a small probable error in no way necessarily guarantees the representativeness of the sample. It indicates only the adequacy of the sample to take care of chance variations, which would be present whether the sample is representative or not. If, however, the measures computed from successive samples chosen in a similar manner show a considerable fluctuation—if we find repeatedly that the averages or other measures fluctuate by more than three times the Standard Error, as computed above, it is a *warning that other than chance factors are causing variations.* The entire procedure of sampling should then be reviewed. It is in this indirect way, in furnishing an index to the probable reliability of the sample and the degree of significance to be attached to the results, that the measures of probability are of great importance in sampling.

We may conclude, then, with reference to the absolute size of the sample:

1. A large sample is more likely than a small sample to include all the significant characteristics of the universe in question if the method of sampling has been random in both cases.
2. The stability of measures computed from a sample increases as the square root of the number of items composing it.
3. The lower limit of the sample is the smallest number which includes all the essential elements.[20]
4. The upper limit is the largest number that can be efficiently or conveniently handled with the available resources.[21]
5. A sample of large size or small probable error, or both, does not necessarily establish the representativeness of the sample.

[19] R. E. Chaddock, *op. cit.*, p. 245.
[20] *Cf.* F. S. Chapin, *op. cit.*, p. 122.
[21] *Ibid.*

E. PRACTICAL TESTS OF THE ADEQUACY OF A SAMPLE

The problem of the size and representativeness of the sample is obviously of great practical concern to any investigator. On the one hand, we wish to generalize from our analysis of the sample, and wish to be as sure as possible that what we find true of the sample would hold true of the whole group from which it is chosen. On the other hand, we wish to expend the least time, money, and effort consistent with the attainment of our result. As we have seen above, when the composition of our whole group is known the minimum size of the sample may be mathematically determined. But where this composition is not known, which is frequently the case in any investigation, and always true in the case of our smallest sub-sample, we must satisfy ourselves of the adequacy and representativeness of our sample by empirical methods.

Perhaps the simplest and most practical empirical method of testing the adequacy of a sample is to take several other samples of about the same size, by the same or a variety of methods, and compare them with the sample selected for investigation. If the results are not sufficiently alike or stable to accord with the standard of accuracy previously determined, it indicates that the size of the sample should be increased until the successive samples show substantial similarity, that is, until most of the variations in each sample are small and occur as frequently in one direction as in the other.[22] The degree of similarity of successive samples which we insist upon will be determined, of course, by the standard of accuracy we have adopted for our investigation.

F. GENERAL CAUTIONS IN SAMPLING

Bowley has well said that "no formal rules can replace judgment and experience in the selection and interpretation of samples."[23] All that rules can do is to cause the investigator to be wary of the more common and obvious errors, and to help him avoid extravagant generalizations from his findings when the validity of the sample is doubtful. Broad generalizations from inadequate and non-representative data

[22] *Ibid.*, p. 125.
[23] A. L. Bowley, *op. cit.*, p. 62.

drawn from personal experience are perhaps one of the commonest sources of error in current social thinking. The public press is full of reports of investigations proving diametrically opposite conclusions, and frequently the explanation is to be found in the methods of sampling employed. Frequently, of course, the mis-selection of the sample is deliberate in order to prove a case for advertising or other propagandist purposes. But often ignorance of the principles of valid sampling is the cause.

As Stephan has said: "In some situations it may be possible to choose a sample which is superior to a random sample. The advantages of deliberately selecting items for a sample, as compared with selecting at random within relevant categories or strata, probably cannot be decided in general terms. Deliberate selection is preferable if the investigator has reasonably complete knowledge about each item. In practical situations, however, he can seldom be sure that he will not unwittingly introduce serious biases in choosing the items. Expediency may dictate the selection of certain cases in the sample. There is then nothing for the investigator to do except to choose as wisely as he can and to keep a wary eye open for clues as to the possible biases that result. The problem focuses on the nature of the social or economic system in which the sampling is conducted, the kind of relations or regressions between relevant factors, and the extent to which the investigator knows about them. With complete knowledge he can pick a perfectly representative sample, with complete ignorance he cannot avoid picking a random one, but with a little knowledge he may be badly misled."[24] Many defects in sampling are due to the common human tendency to select the data which are most convenient and available.[25] For example, Bowley has shown the possible error in taking

[24] F. F. Stephan, "Practical Problems of Sampling Procedure," *op. cit.*, pp. 579-80.

[25] For an excellent example of defective sampling and unwarranted conclusions drawn from the sample, see "The Extent of Old Age Dependency," published by the National Civic Federation (Industrial Welfare Department). See also I. M. Rubinow, "Where Will You Be At 65?" (*New Republic*, 54:289-91, 1928) for an excellent criticism of this study. The sample consists of about one-fourth of one percent of the aged population of the country, all collected in eleven cities in four states on the Atlantic seaboard. It neglects entirely all aged persons already in institutions and contains a disproportionate number of persons of middle-class occupations.

a sample of school children in attendance on a given day if our purpose is to determine the physical condition of school children.[26] Chaddock has likewise illustrated the error possible in a sample of factory workers' family budgets when the accounting required is such that only the members of a certain select group would probably be able and willing to furnish an account of their budgets.[27]

We have already mentioned the biases that may result from geographic inaccessibility, absence of the people to be interviewed, their illness, their inability to speak English, etc. Then there is always the possible bias in samples because some informants refuse to cooperate. If these people do not differ from the rest in *any other way* than in their refusal, there is no bias. But refusal may be associated with characteristics highly important to the investigation: age, education, income. If so, omitting such cases may cause a serious bias. Samples selected by telephone interviews have been shown to be biased against the lower income groups.[28] The selective bias in questionnaires sent out broadcast, and returned in only small proportions is subject to this error. Roper reports that in public opinion polls, at least, "older people answer mail ballots more readily than young people; that the prosperous and well-to-do return a great many more ballots than the poor people, and that the people with violent emotions, particularly with violent emotions *against* whatever is going on, return mail ballots in greater percentage than do the people of more moderate sentiments. Therefore, even though the *Literary Digest* was able in some way to get its ballots to a true cross-section of the public, it was sure that a true cross-section of the public was not returning the ballots. Their returns were heavily weighted with the old who were less strongly pro-Roosevelt than the young, heavily weighted with the prosperous who were more strongly pro-Landon than the poor, and also weighted with a large return from violent critics of the administration who, feeling

[26] *Ibid.*, p. 57.
[27] Chaddock, *op. cit.*, p. 386.
[28] P. Lazarsfeld, *Radio and the Printed Page,* (Duell, Sloan and Pearce, 1940), pp. 13, 32 ff. See also, by the same author, "Interchangeability of Indices in the Measurement of Economic Influences," *Jour. Applied Psychol.,* 23:33-45, 1939.

themselves unrepresented, wanted to cast a mail ballot or write a letter to the editor, or to do something to show their disapproval." [29]

It is impossible to give a list of cautions against all the possible errors of sampling that may occur even in the more conventional type of survey or investigation. Knowledge of rules is no substitute for intelligent alertness. Each investigation may be in some respects unique. For example, interviews with the inhabitants of every tenth house along the streets of a city *may* give a representative sample regarding certain characteristics. But, if the streets are divided into blocks by cross-streets at every tenth house, so that an undue proportion of corner houses are included in the sample, the results may be far from representative. Corner houses may belong to families of higher economic status, or such houses may be shops with families living upstairs. To avoid such difficulties all house numbers for the areas to be sampled might be placed on cards, shuffled, and drawn after the manner of lottery sampling. Or, sample blocks or segments of blocks may be selected on the basis of other known information about the area.[30] By defining a block segment as one including every house from a specified house number to (but not including) another specified number, houses not listed in directories would have an equal chance to be included.

In our discussion of sampling we have assumed for the most part that the practical questions to be answered would specify, at least implicitly, what universe is to be sampled. It is desirable to be explicit on this point because a conclusion may be valid for the sample in some respects but not in any or all other respects. A given sample may represent several different universes, although it represents one or some of them much better than others. Stephan illustrates the matter admirably as follows:

"Questionnaire returns from a group of office workers who are employed by a public utility corporation may be regarded as representing (1) those office workers in the corporation who are disposed to answer questionnaires, (2) all office work-

[29] Elmo Roper, "Sampling Public Opinion," *Jour. Amer. Statis. Asso.*, 35:332, (Part I), 1940.

[30] F. F. Stephan, *op. cit.*, pp. 573, 577.

ers employed by utilities in the same city at the time the sample was taken, (3) all office workers employed by *this* corporation during a period, say five years, preceding (or following) the time the sample was drawn, (4) all office workers in this city, (5) all 'white collar workers,' (6) all utility employees, (7) all residents of the city or area covered by the sample, (8) all residents of similar cities, or any one of many other universes which may be constructed or defined for the purposes of drawing general conclusions from the sample. The entire socio-economic aggregate may be sliced in a multitude of directions, and the sampling relationship may be assumed to exist between a sample and any appropriate slice. Some of these possible universes are quite similar, and investigators frequently shift their discussion or conclusions from one universe to another without justifying the change of reference or satisfying themselves that the sample is a relatively unbiased representative of the second universe as well as the first. This is a weakness common to non-statistical studies quite as much as to statistical investigations. In the case of the latter, measures are available with which each universe may be specified in precise terms." [31]

G. SUMMARY

In conclusion we may summarize the steps of the sampling procedure as follows: [32]

1. Choice of the universe or population to be sampled. All voters or all adults? Native white families only? etc.
2. Decision on a suitable means of access to the population to be sampled. House-to-house canvass? By mail? etc.
3. Choice of a unit of sampling. Individuals or households? Block, census tract, ward, or township? etc.
4. Choice of a method of selecting the sample. Selection by persons very familiar with the population to be sampled? Division of the population into strata known to exist, (*e.g.*, age, sex, economic status) relevant to the inquiry? Various random methods discussed above?
5. Actual selection of the sample. "Drawing" of numbers, selecting houses to be visited, checking sample secured against reliable information already known to detect inadequacies and

[31] F. F. Stephan, *op. cit.*, 574-75.

[32] Adapted from F. F. Stephan, "Representative Sampling in Large Scale Surveys," *Jour. Amer. Statis. Asso.*, 34:347-50, 1939.

biases, adjustments to be made in final results as partial corrections of these biases, etc.

6. Securing data for the units in the sample. Observe, interview, or write to individuals, or transcribe data from records selected by sample.

7. Tabulation and analysis of sample data.

8. Application of the sample data to the problem under consideration. Considerations of reliability, size of errors due to "chance" variation, general adequacy and validity of the data.

We have limited discussion in this chapter only to the more elementary and practical problems of certain types of sampling, especially sampling of populations. But we may also be interested in sampling time, space, or elements constituting cultural concepts. For example, when our data comprise a period of time or even a date, we are really dealing with a sample of all time. It is not a random sample and hence the logic of ordinary standard errors does not apply. It is well to keep in mind, however, in our conclusions, that our inferences regarding trends are based on a sample of time, and any projection of these trends into the future will depend on the adequacy and the validity of our time sample. Any region studied is likewise only a part of a larger region. Finally there is the sampling involved in the construction of the tests and scales (discussed in a later chapter). Thus Chapin, in the construction of his scale[33] for measuring socio-economic status, selects four general components (cultural equipment, effective income, material possessions, and social participation) of status from the whole complex of culture, and then selects a small number of specific items as samples of these four components. S. C. Dodd,[34] likewise, in his construction of a scale for measuring rural hygiene in Syria first formulated some 2000 questions regarding hygiene, and gradually reduced their number by selecting those which correlated highly with the full scale.

These direct techniques of situation sampling of culture and behavior are just as important as the more familiar sampling of people. To be sure, our sampling of people in the social sciences is usually on the basis of the assumption that

[33] F. S. Chapin, *Contemporary American Institutions,* (Harper, 1935), Chap. 19. See also Chap. IX of the present text.

[34] S. C. Dodd, *A Controlled Experiment on Rural Hygiene in Syria* (Oxford Univ. Press, 1934).

they are the carriers of the culture or behavior studied. But in some types of investigation and experimentation we are interested not only in the representative nature of the population, as such. We are also interested in the behavior samples secured from that population. For example, we may observe during definite intervals of time, the occurrence or non-occurrence of certain specified behavior in a group of children. The question of how representative and adequate are the time samples we have taken as compared with what constant observation of the same group would reveal, is here obviously a problem in time sampling.[35] Likewise, in a study of traffic problems we are interested not only in the representativeness and adequacy of the city or the part of the city where the observations are made, but also in the validity of the time samples during which observations were made.

Practical necessities frequently compel us to proceed in investigations when we know that our sampling methods are far from ideal and that our sample is therefore probably defective. Provided we do not make rash generalizations from such studies there is, of course, no objection to them. On the contrary, they may be of great suggestive value and furnish hypotheses for further investigation. A compilation of typical examples of sampling employed in notable research projects would have great suggestive value, as would a compilation of examples of defective sampling. Each new situation, however, may require a new type of sampling. Resourcefulness, imagination, and critical insight are always necessary in selecting a representative sample.

H. CONCLUSION

We have pointed out in this chapter some of the ways in which it is possible to secure reliable knowledge of a vast universe by acquainting oneself with a sample of it. As we have seen, the validity of our assumptions about the whole from a study of a part, depends upon the *representa-*

[35] For further discussion of such techniques, see W. C. Olson and Elizabeth Cunningham, "Time Sampling Techniques," *Child Development*, 5:41-58, 1934. Florence Goodenough, "Measuring Behavior Traits by Means of Repeated Short Samples," *Jour. Juvenile Research*, 12:230-35, 1928. Ruth E. Arrington, "An Important Implication of Time Sampling in Observational Studies of Behavior," *Amer. Jour. Sociol.*, 43:284-95, 1937.

tiveness of that part. The technique of selecting a sample so that there is high probability that it will be representative has been our chief concern.

The dependency of men in all walks of daily life upon the validity of their generalizations from samples is obvious. Were it not for the validity of our assumption that what we have found to hold true in one situation will also be true in other situations which we consider similar, our reactions would be chaotic, and would consist of blind trial-and-error gropings. We learn in actual life to rely upon the theory of sampling long before we formally become aware of its existence. Unfortunately, the simple, primary-group, rule-of-thumb sampling methods that everybody uses from childhood frequently mislead us when we attempt to project them into more remote fields of secondary-group adjustment. The reason is, of course, that in primary group relationships we know a good deal about the universe we are sampling. The opinions of two or three people in our own community are promptly evaluated from the standpoint of the representativeness of these persons—the bias they represent, and the degree to which they probably represent community sentiment. The opinion of the preacher and the banker as well as that of the village "bum" is promptly evaluated from the standpoint of bias and representativeness. In a strange city or in a foreign country, however, we are prone to accept the opinion of any taxi-driver as reliable information about the social, economic, and political conditions in the whole country. If our informant is a prominent person, whether in poetry, chemistry, or journalism, we forget even more completely the elementary rules of valid sampling.

This naïve projection of local primary group sampling methods into secondary group fields is exemplified also by the politicians, journalists, and tourists who are unable and unaccustomed to interpret sources of comprehensive information about a foreign country, and hence decide to spend two or three weeks there to "get the real facts." The incidents which they experience and report, no matter how unique and no matter how small a minority is affected by them, become the stereotypes on the basis of which whole nations are condemned as "barbarian," "godless," or what

not, or, conversely, are glorified as "kindly," "generous," "honest," and "democratic." These illusions frequently contribute not a little to misunderstanding, social tension, and war. The "cases" of social agencies and especially the characters and situations portrayed in novels and the movies are widely accepted as representative. The novel and the play are, in fact, felt by many to be the ultimate source of information about social and psychological matters. *They may or they may not be.* It depends on such criteria as we have discussed in this chapter, not upon whether the tale "rings true." It may ring true merely to our previously formed biases, and if so, our illusions are only confirmed by their corroboration from "literature." Much has been written about the mischief which stereotypes play in human relations. Perhaps one of the basic attacks upon this problem is a more general diffusion of knowledge of the elements of valid sampling techniques.

I. SUGGESTIONS FOR FURTHER STUDY

1. The best brief supplementary reading on the aspects of sampling treated in this chapter will be found in G. U. Yule and M. G. Kendall, *An Introduction to the Theory of Statistics,* (Chas. Griffin and Company, London, 1937), Chap. 18 on "Preliminary Notions on Sampling." Detailed treatment of the statistical-mathematical aspects of large and small samples will also be found in Chaps. 19-23 of this excellent book. The statistics of small samples is perhaps most adequately treated in R. A. Fisher, *Statistical Methods for Research Workers,* Oliver and Boyd, (Edinburgh: 1936). See also his *The Design of Experiments,* Oliver and Boyd, (Edinburgh: 1937). For a more elementary treatment of the statistics of sampling, see F. E. Croxton and D. J. Cowden, *Applied General Statistics,* (Prentice-Hall, 1940), Chap. 12. Also R. E. Chaddock, *Principles and Methods of Statistics,* (Houghton Mifflin, 1925), Chap. 11.

2. An interesting brief description of the sampling technique in the United States Census of 1940 will be found in F. F. Stephan, W. E. Deming, and Morris H. Hansen, "The

Sampling Procedure in the 1940 Population Census," *Jour. Amer. Statis. Asso.,* 35:615-30, 1940. See also F. F. Stephan and W. E. Deming, *On the Sampling Methods in the* 1940 *Population Census,* Bureau of the Census, 1941.

3. Further details on the sampling techniques of the principal public opinion polls will be found in George Gallup and S. F. Rae, *The Pulse of Democracy,* (Simon and Schuster, 1940). A briefer treatment of the same subject is Elmo Roper, "Sampling Public Opinion," *Jour. Amer. Statis. Asso.,* 35:325-34, 1940.

4. Descriptions of the sampling problems encountered and the techniques adopted in concrete investigations of different kinds will be found in the following sources:

(*a*) Erika H. Schoenberg and Mildred Parten, "Methods and Problems of Sampling Presented by the Urban Study of Consumer Purchases," *Jour. Amer. Statis. Asso.,* 32:311-22, 1937.

(*b*) The *National Health Survey, Preliminary Reports: Significance, Scope and Method,* (Washington: U. S. Public Health Service, 1938).

(*c*) Calvert L. Dedrick and M. H. Hansen, "The Enumerative Check Census," *Census of Partial Employment, Unemployment and Occupations; 1937.* (Washington, 1938), pp. 159 ff.

(*d*) C. F. Sarle, "Development of Partial and Sample Census Methods," *Jour. Farm Economics,* 21:357-64, 1939; Irvin Holmes, "Results of Four Methods of Sampling Individual Farms," *Ibid.,* 365-74.

(*e*) Paul F. Lazarsfeld, *The American Station Sampler and How to Use It,* (Office of Radio Research, Columbia University, 1940).

5. Excellent supplementary reading on the general and practical aspects of sampling will be found in two articles by F. F. Stephan, "Representative Sampling in Large-Scale Surveys," *Jour. Amer. Statis. Asso.,* 34:343-52, 1939; and "Practical Problems of Sampling Procedure," *Amer. Sociol. Review,* 1:569-80, 1936. See also S. A. Stouffer, "Sociology and Sampling," *Fields and Methods of Sociology,* Edited by L. L. Bernard, (Long and Smith, 1934), Chap. 15.

CHAPTER VI

THE SCHEDULE AS AN INSTRUMENT OF OBSERVATION

The next care to be taken in respect of the Senses, is a supplying of their infirmities with *Instruments*, and, as it were, the adding of *artificial Organs* to the natural; this in one of them has been of late years accomplished with prodigious benefit to all sorts of useful knowledge, by the invention of Optical Glasses. . . By this the Earth itself, which lyes so neer us, under our feet, shews quite a new thing to us, and in every *little particle* of its matter, we now behold almost as great a variety of Creatures, as we were able before to reckon up in the whole Universe itself. It seems not improbable, but that by these helps the subtility of the composition of Bodies, the structure of their parts, the various texture of their matter, the instruments and manner of their inward motions, and all the other appearances of things, may come to be more fully discovered. — *Robert Hooke.*[1]

A. THE NEED OF INSTRUMENTS OF OBSERVATION IN THE SOCIAL SCIENCES

Man is usually credited with being the only tool-making and tool-using animal. To this proclivity he owes his domination over other animals and over his environment generally. Man's first concern with tools was perhaps with their usefulness as extensions of his more overt muscular powers. But he soon developed rude devices to guide and extend his observations, and to check them objectively. When the savage put his ear to the ground the better to detect sounds, or shaded his eyes with his hand to aid his vision, he made the first rude gestures in the extension of his stimuli receptors, the highest form of which we now find in our most powerful microscopes, telescopes, and various delicate instruments of sound detection. When he gauged the time of day by the position of the sun relative to a tree, the passage of longer periods of time by the phases of the moon, and distance in terms of a day's march, he was beginning to check his subjective impressions by objective criteria. Man's progress toward more delicate and accurate adjustments has consisted of the increasing refinement of these mechanical devices for extending and objectifying his observations. In physical

[1] *Micrographia*, 1665. (Preface, pp. 8, 9, Alembic Club Reprints. William F. Clay, London and Edinburgh, 1894.)

science that refinement has reached the highest point yet attained in any field.

The importance of supplementing and extending the power of our stimuli receptors or "senses" by mechanical devices was stressed in a previous chapter. No fact in the history of science stands out more clearly than the close dependence of scientific advancement on the invention of mechanical aids to observation. In the hands of Galileo the telescope revolutionized the then existing notions of the universe. Recent developments in telescopic, photographic, and spectroscopic instruments enable us to make physical and chemical analyses of the heavenly bodies. Numerous phenomena which for centuries puzzled the ablest minds because their instruments of observation were inadequate are now well understood. Without the development of the microscope, bacteriology could not have developed at all and other biological sciences would have been seriously handicapped. Many overlook how completely the imposing edifice of physical science rests upon the intricate and ingenious mechanical devices we have invented to extend and objectify our powers of observation. Continued advancement in our knowledge, furthermore, must depend very largely upon the increasing refinement of these devices.

As was also pointed out in a previous chapter, one of the reasons for the low state of development of the social sciences is the comparative absence of accurate and intensive observation devices. From the rôle which these devices have played in the development of the physical sciences, we may assume that our first task in the social sciences is the improvement of our mechanical aids to observation. Our chief reliance so far has been on some form of a *schedule*. This device is an aid to observation in several ways.

B. THE IMPORTANCE OF THE SCHEDULE AS AN AID TO OBSERVATION

In the first place, the schedule extends enormously the investigator's powers of observation. It does this by reminding the observer of each item which is to be looked for and reported on. The schedule insures that the observer makes note of the same items of information in all the cases, and

yet makes it unnecessary for him to trust his memory for the facts observed. Conclusions based on a large number of observations without the assistance of a formal schedule are notoriously untrustworthy. An inspection of even a small number of industrial establishments, for example, without some such schedule as illustrated on page 170 would enable the investigator to give only a very vague and inaccurate account of his observations. In the first place, he would fail to observe many of the items called for on the card. In the second place, he could not remember the number of cases in which a particular item was present or absent or the degree to which it was present or absent. The investigator would tend to observe and remember only the unique, exceptional, or unusual features, or the aspects in which he was especially interested. The conflicting reports of visitors and "investigators" of Russia in recent years furnish excellent examples of what happens when we attempt to carry on observations without the aid of some objectifying device. Visitors to the same place at the same time have brought back radically different reports of what they saw. The notions of almost any community on changes in climate, on the comparative severity of winters, etc., are examples of the same error in generalizations which are not based on objective recording of the facts. A check of the records of the Weather Bureau will show in most cases that general impressions are not supported by the facts. It is in the elimination of such errors that the schedule serves as an important tool in extending and validating the observation of social facts.

In the second place, the schedule tends to standardize and objectify the observations of different observers. That the facts observed shall appear the same to all qualified observers, is, as we have seen, one of the primary requirements of science. The schedule helps to achieve this result (1) by singling out the particular aspects of a situation which are regarded as significant for the purpose in hand and calling attention to these; and (2) by specifying in advance the units and terminology in which the observations are to be described. As we have seen in a previous chapter, in order to make a unit mean the same to different persons we objectify it in some way by referring it to some external measuring

device, such as a scale, a yardstick, etc. That is, we tend to describe our observations quantitatively instead of qualitatively. The importance of quantitative description in science has been emphasized in previous chapters.

Finally, the schedule is a device for isolating one element at a time and thus intensifying our observation of it. This again is one of the techniques on which physical science has depended for much of its advancement. The division of a mass of impressions into its constituent factors is the essence of all analysis, and only through such a separation of factors can sequences be observed and behavior understood.

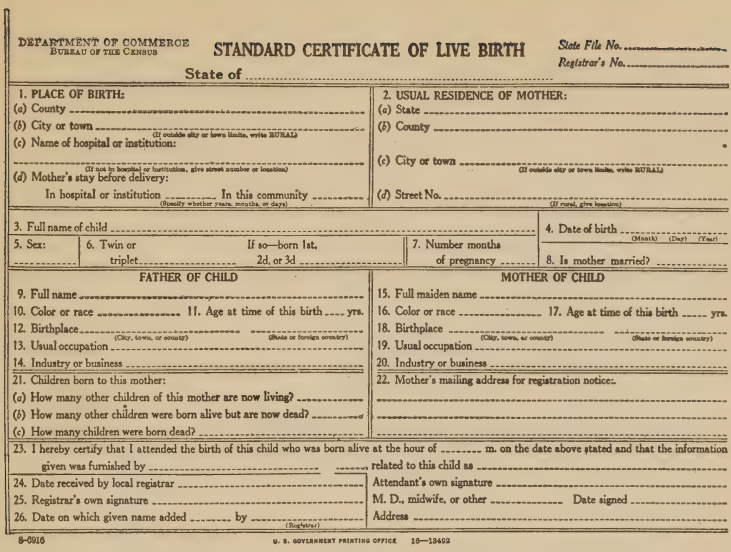
The careful construction and use of schedules, therefore, becomes one of our most important problems in the observation of social phenomena.

C. TYPES OF SCHEDULES

For convenience in discussion we shall divide schedules into three classes: (1) Schedules for the recording of objective facts (*a*) from the direct observation of the investigator, (*e.g.*, Fig. 2), (*b*) from interviews with people who possess the desired facts (*e.g.*, a birth registration schedule, Fig. 1, or the United States Census population schedule, Fig. 2), and (*c*) from "questionnaires" filled out by people without personal interview by an investigator; (2) schedules for the determination and measurement of attitudes or opinions, (see Chap. VIII) and (3) schedules for the scoring of the status and functioning of social organizations and institutions so that they may be compared or measured relative to some standard and relative to each other (see Chap. IX). The above categories are not mutually exclusive, of course. Many schedules would fall into two or more of the classes given. The present chapter will deal with only the first of these classes.

D. SCOPE AND PURPOSE OF THE INVESTIGATION

Before schedules can be intelligently prepared, the scope, plan, and purpose of the investigation must be decided upon. Some of the considerations involved in this preliminary work were discussed in the preceding chapters in connection with the discussion of the working hypothesis and sampling. We pointed out in Chapter I that too many projects simply un-



DEPARTMENT OF COMMERCE
BUREAU OF THE CENSUS

STANDARD CERTIFICATE OF LIVE BIRTH

State of .....

State File No. .....
Registrar's No. .....

1. PLACE OF BIRTH:
(a) County .....
(b) City or town ..... (If outside city or town limits, write RURAL)
(c) Name of hospital or institution:
(If not in hospital or institution, give street number or location)
(d) Mother's stay before delivery:
In hospital or institution ..... In this community .....
(Specify whether years, months, or days)

2. USUAL RESIDENCE OF MOTHER:
(a) State .....
(b) County .....
(c) City or town ..... (If outside city or town limits, write RURAL)
(d) Street No. ..... (If rural, give location)

3. Full name of child .....
4. Date of birth ..... (Month) (Day) (Year)
5. Sex: .....
6. Twin or triplet ..... If so—born 1st, 2d, or 3d .....
7. Number months of pregnancy .....
8. Is mother married? .....

FATHER OF CHILD

9. Full name .....
10. Color or race ..... 11. Age at time of this birth ..... yrs.
12. Birthplace ..... (City, town, or county) ..... (State or foreign country)
13. Usual occupation .....
14. Industry or business .....
21. Children born to this mother:
(a) How many other children of this mother are now living? .....
(b) How many other children were born alive but are now dead? .....
(c) How many children were born dead? .....

MOTHER OF CHILD

15. Full maiden name .....
16. Color or race ..... 17. Age at time of this birth ..... yrs.
18. Birthplace ..... (City, town, or county) ..... (State or foreign country)
19. Usual occupation .....
20. Industry or business .....
22. Mother's mailing address for registration notice: .....

23. I hereby certify that I attended the birth of this child who was born alive at the hour of ..... m. on the date above stated and that the information given was furnished by .... related to this child as .....
24. Date received by local registrar .....
25. Registrar's own signature .....
26. Date on which given name added ..... by ..... (Registrar)

Attendant's own signature .....
M. D., midwife, or other ..... Date signed .....
Address .....

8-0916 U. S. GOVERNMENT PRINTING OFFICE 16—13492

WHY BIRTHS SHOULD BE REGISTERED

There is hardly a relation of life, social, legal, or economic, in which the evidence furnished by an accurate registration of births may not prove to be of the greatest value, not only to the individual but also to the public at large. It is not only an act of civilization to register birth certificates but good business, for they are frequently used in many practical ways:

(1) As evidence to prove the age and legitimacy of heirs;
(2) As proof of age to determine the validity of a contract entered into by an alleged minor;
(3) As evidence to establish age and proof of citizenship and descent in order to vote;
(4) As evidence to establish the right of admission to the professions and to many public offices;
(5) As evidence of legal age to marry;
(6) As evidence to prove the claims of widows and orphans under the widows' and orphans' pension law;
(7) As evidence to determine the liability of parents for the debts of a minor;
(8) As evidence in the administration of estates, the settlement of insurance and pensions;
(9) As evidence to prove the irresponsibility of children under legal age for crime and misdemeanor, and various other matters in the criminal code;
(10) As evidence in the enforcement of law relating to education and to child labor;
(11) As evidence to determine the relations of guardians and wards;
(12) As proof of citizenship in order to obtain a passport;
(13) As evidence in the claim for exemption from or the right to jury and military service.

FIGURE 1. A Simple Type of Schedule.

dertake to study "unemployment," "women in industry," "relief," "crime," etc., without specifying clearly the exact aspects of these phenomena to be studied. Accordingly, schedules are hastily made up to consist of any or all questions occurring to anyone as being "interesting." Too many schedules are mere inventories of what someone would like to know. They fail to take into consideration the problem of what we may reasonably hope to get, and the best way of getting it. Inexperienced investigators are especially prone to rush into the field before the purpose of the research has been carefully thought through. Their schedules will be correspondingly vague and ambiguous. Social and economic *subjects* have an indefinitely large number of aspects which may be investigated. A choice must be made. We may set it down as a first rule not to attempt the preparation of a schedule for an investigation until the purposes of the inquiry have been reduced to *specific questions*. This does not mean that *general information* may not be gathered on as general subjects as is desired. It merely means that *whatever is desired should be explicitly stated.*

When it has been definitely decided what questions are needed, the next step is a careful consideration of just what data, when gathered and analyzed, would yield the answers sought. The availability of these data within the time and other resources of the investigation must then be considered. It may be known from the outset: (*a*) that some of the required data are certainly available, (*b*) that certain other necessary data definitely are not available, and (*c*) that other data are of doubtful availability. If the questions call for unavailable data, this fact had better be recognized in advance. Otherwise the time of the investigators will be wasted on a vain quest. When availability is in doubt, the point should be cleared up as far as possible by preliminary inquiry. If it is certain the data will not be available to settle the questions as originally propounded, it may be possible to rephrase them so that the original purposes can be achieved in whole or part through indirect or inferential channels. Historians and anthropologists, not to mention students of cosmic and terrestrial evolution, frequently cannot hope to find direct or concrete accounts of many of the events of the indefinite

past. But through study of contemporary data they can make fairly reliable deductions regarding what did occur, because of what we know to be probable under certain circumstances. Data from which such inferences can be drawn are, of course, just as proper objects of inquiry as direct census information about birth rates, death rates, or unemployment. Except in purely exploratory investigations it serves no purpose to proceed with an investigation until the availability of the data has been determined with some certainty. Hypotheses couched in such terms as to be unsusceptible of empirical test are metaphysical rather than scientific hypotheses.

Secrist[2] has summarized this preliminary work in the form of nine questions which should be answered before the collection of data begins. While these points contemplate a statistical investigation, they would apply with equal validity to any other form of inquiry:

1. What is the precise problem upon which statistics are required?
2. Does the problem as formulated lend itself to statistical treatment?
3. What types of data are necessary for its analysis or solution?
4. Are they likely to be available in suitable form?
5. Are they likely to be adequate for the purpose in mind?
6. Will they have the required degree of accuracy, consistency, and comparability?
7. Can the data be made available within the time limit required; that is, have they the required currency?
8. Are there likely to be any restrictions upon the use of data which will compromise the purpose which they are to serve?
9. What sanction is necessary, and what method of procedure, with the sanction available, must be followed in order to secure the desired facts?

E. THE CONSTRUCTION OF SCHEDULES

Since the value of the results of an investigation will frequently be determined by the care with which the schedule is constructed, it is advisable to prepare it with great care. This is especially true in the case of the questionnaire. Some special considerations concerning this type of schedule will be

[2] Horace Secrist, *An Introduction to Statistical Methods*, (Revised edition, Macmillan, 1925), p. 47.

considered in the next chapter. But even the enumeration schedule deserves careful construction from the standpoint of uniformity and reliability of the results it will bring in, as well as from the standpoint of clearness to tabulators and interpreters of the results. The principal aspects of schedule making that demand attention, and which apply to all the forms mentioned above, may be summarized as follows:

1. *Items to include*

To determine what items to include in a schedule, a careful preliminary study of the problem is necessary. The temptation is to include all items which come to mind as being "interesting" or "possibly of value." But this tendency has to be avoided because in a comprehensive study, a single additional item may cost thousands of dollars in the additional amount of time and space required for recording and the extra work involved in editing and tabulating. The principal rule should be: *Never include an item without first having a clear idea of exactly how the information is to be utilized and what it will contribute to the purpose sought in the investigation.* If its importance from this point of view is low compared with the additional cost of including it, the item should be discarded.

There are, however, questions and items which, although they might be of considerable value, arouse antagonism on the part of the informant. These it may be wise to omit. This is especially true when the investigator is dependent on the cooperation of the informant for his information. Many people are inclined to be reticent on questions of mental or physical infirmities, marital and sexual relations, the use of intoxicants, etc., and attempts to secure information on such matters will usually arouse antagonism. Unless the purpose of the investigation distinctly calls for information on these topics, they had better be avoided, for indelicate inquiries on them may seriously jeopardize the more important information we want to secure. These problems may be met in part by careful consideration of the wording of the inquiry.

2. *Dummy Tables*

One of the best ways of making sure that the schedule

will call for the pertinent data required by the investigation is to prepare in advance imaginary dummy or "ghost" tables. A dummy table is one which, if properly filled in, would answer the questions we want answered. With these skeleton tables before us, the next question is, what specific inquiries in the schedule would yield data of the type and in the *form* necessary to construct such tables? By constantly keeping in mind the desired result of the investigation, most of the problems of inclusion or exclusion can be settled on this practical basis.

3. *Wording*

Headings and titles should be sufficiently complete to indicate fully the kind of data contained under them, and sufficiently clear to be understood by a person of ordinary intelligence. Words and phrases of ambiguous meaning should be avoided. The enumerator, at least, should understand exactly what interpretation each item calls for. For example, if the item is "age," this is subject to at least three interpretations: (1) exact present age, (2) age at last birthday, (3) age at nearest birthday. Accuracy and uniformity demand that this item be defined in the same way in every case, and the schedule should make the interpretation clear. When the schedules are to be filled out by skilled field workers, they may be given uniform instructions and interpretations in advance. This will make unnecessary such rigid specifications in the schedules as is necessary when they are to be used by unskilled workers. In some cases the necessary interpretations may be so complex as to require more extensive instructions than can be printed on the face of the schedule. Thus the Federal Census enumerators are equipped with a separate manual of instructions that aims to cover in full all questions of interpretation of each item. Consider, for example, the following extracts from the instructions to enumerators in connection with the recording of occupations:

544. The term "laborer" should be avoided if any more precise statement of the occupation can be secured. Employees in factories and mills, for example, usually have some definite designation, as weaver, roller, etc. Where the term "laborer" is used, be especially careful to state accurately the industry or business in col. 29.

545. Avoid the use of the word "mechanic" whenever a more specific occupation can be given, such as *carpenter, painter, electrician,* etc. *Automobile mechanic,* however, is a satisfactory return.

546. Avoid the use of the word "clerk" wherever a more definite occupation can be named. Thus, an employee in a store who is wholly or principally engaged in selling goods should be called a *salesman* and not a "clerk." A *typist, accountant, bookkeeper, cashier,* etc., should be reported as such, and not as a "clerk." Do not return a *stenographer* as a "secretary." Distinguish a *traveling* salesman from a *salesman* in a store. . .

550. For an employee who works for a concern that carries on different activities, the return should be the industry in which he directly works, provided that represents a major subdivision of the enterprise and is carried on in a separate building or in a place physically distinct from other activities of the concern. Thus, a miner working in a coal mine, owned and operated by a steel mill, should be returned as engaged in the coal mining industry (that is, in his immediate place of work) and not as working for a steel mill. On the other hand, persons working in a department or other unit incidental to the main work of an establishment should be returned in the Industry column (col. 29) as engaged in the major activity of the establishment. For example, the industry return for workers in the power plant of a steel rolling mill should be *steel rolling mill;* that for workers in the box making department of an electrical machinery factory should be *electrical machinery factory;* that for workers in the garage or the warehouse of a department store should be *department store;* and that for workers in the foundry of an agricultural implement factory should be *agricultural implement factory. . .*

552. Avoid General or Indefinite Terms. — Give the occupation and industry precisely. For example, return a worker in a textile mill as a spinner, *cotton mill;* weigher, *woolen mill;* spooler, *silk mill;* etc. Do not report industry in such indefinite terms as refinery, transportation, electrical, etc., but specify *petroleum refinery, sugar refinery,* or *copper refinery; steam railroad,* or *bus line; electrical power company* or *electrical appliance factory,* etc. Never enter in col. 29 such indefinite terms as "factory," "mill," "shop," "store," or "office," without stating the kind of factory, etc., as *soap factory, cotton mill, auto repair shop, grocery store, real estate office.*[3]

Whenever possible it is advisable to have the wording of the schedule itself make clear the exact meaning of the item.

[3] Sixteenth Census of the United States, 1940, *Instructions to Enumerators,* pp. 61-63.

ILLUSTRATIVE EXAMPLE

"Your report is made of by Act of Congress. This Act makes it unlawful for the Bureau to disclose any facts. Individual names or identity from your answers supplied will use your statements from collected will be used solely for preparing statistical information concerning the Nation's population, resources, and business activities. Your Census Reports Cannot Be Used for Purposes of Taxation, Regulation, or Investigation

State Iowa — Incorporated place Orange city — Ward of city 2 — Unincorporated place ... (Name of unincorporated place having 100 or more inhabitants)

County Stanton — Township or other division of county Royce — Block Nos. 370-371 — Institution ... (Name of institution and kind, as whist' 'nurses' or 'asylum')

DEPARTMENT OF COMMERCE—BUREAU OF THE CENSUS

SIXTEENTH CENSUS OF THE UNITED STATES: 1940

POPULATION SCHEDULE

E.D. No. 9 — E.D. No. 71-14 — Sheet No. 9 A

Enumerated by me on April 2 1940

Harold W. Thompson — Enumerator.

Line No.	LOCATION: Street, avenue, road, etc.	LOCATION: House number (in cities and towns)	HOUSEHOLD DATA: Number of household or order of visitation	HOUSEHOLD DATA: Home owned (O) or rented (R)	HOUSEHOLD DATA: Value of home, if owned, or monthly rental, if rented	HOUSEHOLD DATA: Does this household live on a farm?	NAME: Name of each person whose usual place of residence on April 1, 1940, was in this household. BE SURE TO INCLUDE: Persons temporarily absent from household. Write 'At' after names of such persons. Children under 1 year of age. Write "Infant" if child has not been given a first name. Enter ⓧ after name of person furnishing information.	RELATION: Relationship of this person to the head of the household, as: wife, daughter, father-in-law, grandson, lodger, boarder, servant, hired hand, etc.	PERSONAL DESCRIPTION: Sex — Male (M), Female (F)	PERSONAL DESCRIPTION: Color or race	PERSONAL DESCRIPTION: Age at last birthday	PERSONAL DESCRIPTION: Single (S), Married (M), Widowed (Wd), or Divorced (D)	EDUCATION: Attended school or college (any time since March 1, 1940)? (Yes or No)	EDUCATION: Highest grade of school completed	PLACE OF BIRTH: If born in the United States, give State, Territory, or possession. If foreign born, give country or name at present. Distinguish Canada-French from Canada-English and Irish Free State (Eire) from Northern Ireland.	CITIZENSHIP: Citizenship of the foreign born	RESIDENCE, APRIL 1, 1935: IN A DIFFERENT PLACE: City, town, or village having 2,500 or more inhabitants. Enter "R" for all other places.	RESIDENCE, APRIL 1, 1935: IN A DIFFERENT PLACE: County	RESIDENCE, APRIL 1, 1935: IN A DIFFERENT PLACE: State or Territory or foreign country	RESIDENCE, APRIL 1, 1935: IN A DIFFERENT PLACE: On a farm? (Yes or No)	PERSONS 14 YEARS OLD AND OVER—EMPLOYMENT STATUS: Whether working for pay or profit during week (or last regular week) (Yes or No)	PERSONS 14 YEARS OLD AND OVER—EMPLOYMENT STATUS: If not working, whether seeking work (Yes or No)	PERSONS 14 YEARS OLD AND OVER—EMPLOYMENT STATUS: If neither working nor seeking work, whether assigned to public emergency work (WPA or Navy Y; CCC or Army C; NYA or Colls 1 to 4; Civil Service CS)	PERSONS 14 YEARS OLD AND OVER—EMPLOYMENT STATUS: For persons working or assigned to public emergency work (Col. 22 or 23): Number of hours worked during week	PERSONS 14 YEARS OLD AND OVER—EMPLOYMENT STATUS: If not at work (Col. 21 blank): Number of weeks worked in 1939	PERSONS 14 YEARS OLD AND OVER—EMPLOYMENT STATUS: If seeking work or assigned to public emergency work (Col. 22 or 23): Number of weeks looking for work in 1940	PERSONS 14 YEARS OLD AND OVER—EMPLOYMENT STATUS: OCCUPATION, INDUSTRY, AND CLASS OF WORKER: For a person at work, assigned to public emergency work, or with a job, "Yes" in Col. 21, 22, or 23: Enter present occupation, industry, and class of worker. For a person seeking work, "No" in Col. 21, 22, or 23: Enter occupation, industry, and class of worker in last job at which he worked 1 day or more since December 31, 1939. If never worked, enter "New worker." If in Col. 21, 22, or 23, and has previous work experience, enter "Have worked" in Col. 24. If never worked, enter "New worker" in Col. 24.	PERSONS 14 YEARS OLD AND OVER—EMPLOYMENT STATUS: OCCUPATION: Trade, profession, or particular kind of work, as: *frame spinner, salesman, laborer, rivet heater, music teacher*	PERSONS 14 YEARS OLD AND OVER—EMPLOYMENT STATUS: INDUSTRY: Industry or business, as: *cotton mill, retail grocery, farm, shipyard, public school*	PERSONS 14 YEARS OLD AND OVER—EMPLOYMENT STATUS: CODE (Leave blank)	PERSONS 14 YEARS OLD AND OVER—EMPLOYMENT STATUS: Class of worker	PERSONS 14 YEARS OLD AND OVER—EMPLOYMENT STATUS: CODE (Leave blank)	INCOME IN 1939: Number of weeks worked in 1939	INCOME IN 1939: Amount of money wages or salary received (dollars)	INCOME IN 1939: Did this person receive income of $50 or more from a business (other than farming), profession, or commercial establishment he owned? (Yes or No)	Number of Farm Schedule	Line No.
1	Rogers Street	317	195	O	4500	No	Owens, Robert S. ⓧ	Head	M	W	56	M	No	C-5	England	Na	Same house				Yes	-	-	-	-	45		Medical doctor	Private practice	OA		52	0	Yes		1
2							——, Ethel	Wife	F	W	52	M	No	C-4	Northern Ireland	Na	Same house				Yes	-	-	-	-	35		Receptionist	Doctor's office	NF		52	0	No		2
3							Lancaster, John J.	Nephew	M	W	14	S	Yes	H-1	Indiana		R	Denton	Indiana	Yes	No	No	Yes	-	-		2	Delivery boy	Retail grocery store	PW		13	130	No		3
4							Moore, Roberta B. ⓧ	Lodger	F	W	27	S	No	C-4	Canada-English	Al	Detroit		Michigan	No	No	No	No	Yes				Art teacher	Private high school	PW		42	1200	No		4
5		196	R	35	No		Riggs, Maude ⓧ	Head	F	W	32	S	No	C-4	Kansas		Lincoln		Nebraska	No		Yes	-	-	-	36		Music teacher	Own studio	OA		52	0	Yes		5
6							Sanders, Grace T.	Partner	F	W	30	S	No	C-4	Illinois		Same house					Yes	-	-	-	42		Librarian	Public library	GW		52	1500	No		6
7							Leeds, Ruth M.	Housekeeper	F	W	26	M	No	8	Iowa		Tipton	City Cedar	Iowa	No		Yes	-	-	-	48		Housekeeper	Private family	PW		52	260	Yes		7
8							——, Alice M.	Housekeeper's daughter	F	W	4	S	No	0	Iowa		-																				8
9		321	197	R	20	No	Anderson, Carl	Head	M	W	37	M	Yes	H-3	Norway	Fa	-	-	Norway	No		No	No	No	Yes	-		Spray painter	Retail auto dealer	PW		40	1400	No		9
10							——, Rita ⓧ	Wife	F	W	35	M	No	H-3	Norway	Fa	-	-	Norway	No		Yes	-	-	-	8		Seamstress	At home	OA		8	0	Yes		10
11							——, Charles	Son	M	W	10	S	Yes	5	Norway	Al	-	-	Norway	No																	11
12							——, Edward	Son	M	W	1	S	No	0	Iowa		-																				12
13		325	198	O	3000	No	Ranson, Harold T.	Head	M	W	52	M	No	H-2	Missouri		Same house					Yes	-	-	-	48		Stationary engineer	Flour mill	PW		42	1425	Yes		13
14							——, Celia ⓧ	Wife	F	W	49	M	No	8	Tennessee		Same house					No	No	No	No	H						4	80	No		14
15							——, Joan Ab	Daughter	F	W	18	S	Yes	G-1	Iowa		Same house					Yes	-	-	-	16		Waitress	School dormitory	PW		19	70	Yes		15
16							——, Thomas B.	Son	M	W	13	S	Yes	H-1	Iowa		Same house																				16
17							——, Betty S.	Daughter	F	W	7	S	No	0	Iowa		Same house																				17
18							Woods, Frank T.	Son-in-law	M	W	24	M	No	C-2	Illinois		Evanston		Illinois	No		No	No	Yes	-	-		19	Attendant	Gasoline filling station	PW		37	740	No		18
19							——, Anne R.	Daughter	F	W	22	M	No	C-1	Missouri		Same house					No	No	No	No	H						0	0	No		19
20							——, Infant	Grandson	M	W	0/12	S	No	0	Iowa		-																				20
21	Austin Avenue	355 (rear)	199	R	10	No	Jackson, Ora	Head	F	Neg	43	Wd	No	3	Louisiana		R	Martin	Michigan	Yes		Yes	-	-	-	60		Maid	Private home	PW		52	520	Yes		21
22							Brown, Sara	Daughter	F	Neg	25	D	No	7	Michigan		Same place					Yes	-	-	-	24		Laundress	Own home	OA		26	0	Yes		22
23																	Here ends Block 370 and begins Block 371																			23
24		648	200	R	30	No	Chalmers, T. Ralph	Head	M	W	30	M	No	H-4	Vermont		Burlington		Vermont	No		Yes	-	-	-	38		Fireman	Newspaper	PW		52	1800	No		24
25							——, Susan ⓧ	Wife	F	W	29	M	No	H-4	New Hampshire		Burlington		Vermont	No		No	No	No	No	H						0	0	No		25
26							——, Alice	Daughter	F	W	7	S	Yes	2	Vermont		Burlington		Vermont	No																	26
27		644	201	R	23	No	Willard, Charles	Head	M	W	39	M	No	H-2	South Carolina		Hoboken		New Jersey	No		Yes	-	-	-	40		Messenger	District WPA office	GW		44	990	No		27
28							——, Ethel ⓧ	Wife	F	W	33	M	No	8	Florida		Hoboken		New Jersey	No		No	No	Yes	-	-		17	New worker				0	0	No		28
29			202	R	23	No	Rodman, George J.	Head	M	W	41	M	No	7	Netherlands	Na	Same house					No	Yes	-	-	-	56		Laborer	Park maintenance project	GW		28	332	No		29
30							——, Catherine P. ⓧ	Wife	F	W	39	M	No	H-2	Illinois		Same house					No	No	No	No	H						0	0	No		30
31							——, Henry Ab	Son	M	W	18	S	Yes	H-3	Iowa		Same house					No	Yes	-	-	-	60		Truck driver	Reforestation project	GW		25	188	No		31
32							Muse, Bertha L.	Mother-in-law	F	W	72	Wd	No	4	Switzerland	Na	R	Adams	Iowa	Yes		No	No	No	No	U						0	0	Yes		32
33							Carter, Evelyn M.	Sister-in-law	F	W	44	D	Yes	8	Iowa		R	Adams	Iowa	Yes		No	No	Yes	-	-		65	Saleswoman	Retail millinery store	PW		0	0	Yes		33
40																																									40

SUPPLEMENTARY QUESTIONS

For Persons Enumerated on Lines 14 and 29

Line No.	NAME	FOR ALL PERSONS: PLACE OF BIRTH OF FATHER AND MOTHER: If born in the United States, give State, Territory, or possession. If foreign born, give country in which birthplace was situated on January 1, 1937 (Distinguish Canada-French from Canada-English and Irish Free State (Eire) from Northern Ireland): FATHER	MOTHER	CODE (Leave blank)	MOTHER TONGUE (OR NATIVE LANGUAGE): Language spoken in home in earliest childhood	CODE (Leave blank)	VETERANS: Is this person a veteran of the United States military forces; or the wife, widow, or under-18-year-old child of a veteran?	VETERANS: If "Yes," enter "1" in the proper column to indicate which war or expedition	VETERANS: World war (1917 or 1918)	CODE (Leave blank)	SOCIAL SECURITY: Does this person have a Social Security account number? (Yes or No)	SOCIAL SECURITY: Did he work in employment or business covered by Social Security during 1939? (Yes or No)	SOCIAL SECURITY: Did he receive wages or salary in 1939 for work in employment covered by Social Security? (Yes or No)	FOR PERSONS 14 YEARS OLD AND OVER: USUAL OCCUPATION, INDUSTRY, AND CLASS OF WORKER: Enter that occupation which the person regards as his usual occupation and at which he is physically able to work. If the person is unable to determine this, enter that occupation at which he has worked longest during the past 10 years and at which he is physically able to work. Enter also usual industry and usual class of worker. If a person without previous work experience, enter "None" in Col. 42 and leave Cols. 43 and 44 blank. USUAL OCCUPATION	USUAL INDUSTRY	Usual class of worker	CODE (Leave blank)	FOR ALL WOMEN WHO ARE OR HAVE BEEN MARRIED: Married more than once? (Yes or No)	Age at first marriage	Number of children ever born (Do not include stillbirths)	Line No.
	35	36	37	0	38	H	39	40	41	I	42	43	44	45	46	47	J	48	49	50			
14	Ranson, Celia	Austria	Austria		German		Yes		W		Yes	Yes	1	Saleswoman	Department store	PW		No	20	5				14
29	Rodman, George J.	Netherlands	Netherlands		Dutch		No				Yes	Yes	3	Laborer	Building industry	PW								29

FOR OFFICE USE ONLY—DO NOT WRITE IN THESE COLUMNS

Tens (K)	Yrs (L)	Fm (M)	Sex (N)	Color and Rel. (O, P, Q, R, S)	Age (T)	Mar (U)	Gr. sch. (V)	Occ (W)	Wrk at (X)	Occupation, industry, and class of worker (Y)	Wks wrk (Z)	Wages (AA)	Ot inc. (BB)	Line No.
K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	
																14
																29

SYMBOLS AND EXPLANATORY NOTES

Dec. 14, 1939 16-26276

Col. 9. VALUE OF HOME, IF OWNED: Where owner's household occupies only a part of a structure, estimate value of portion occupied by owner's household. Then the value of the unit occupied by the owner of a two-family house might be approximately one-half the total value of the structure.

Col. 10. COLOR OR RACE:

White	W	Filipino	Fil
Negro	Neg	Hindu	Hin
Indian	In	Korean	Kor
Chinese	Chi	Other races, spell out in full.	
Japanese	Jp		

Col. 11. AGE AT LAST BIRTHDAY. Enter age of children born on or after April 1, 1939, as follows:

Born in:		Born in:	
April 1939	0/12	October 1932	6-12
May 1939	0/12	November 1938	1/12
June 1939	0/12	December 1938	2/12
July 1939	0/12	January 1939	3/12
August 1939	0/12	February 1939	1/12
September 1939	0/12	March 1939	0/12

(Do not include children born on or after April 1, 1940.)

Col. 14. HIGHEST GRADE OF SCHOOL ATTENDED:

None	0
Elementary school, 1st to 8th grade	1, 2, etc., to 8
High school, 1st to 4th year	H-1, H-2, H-3, H-4
College, 1st to 4th year	C-1, C-2, C-3, C-4
College, 5th or subsequent year	C-5

Col. 16. CITIZENSHIP OF THE FOREIGN BORN:

Naturalized	Na
Having first papers	Fa
Alien	Al
American citizen born abroad	Am Cit

Col. 21. WAS THIS PERSON AT WORK? Enter "Yes" for persons at work for pay or profit in private or Government industry, business, or service during the week (or last regular week) for which hours or days of the hourly wage without money wages or salary at work (other than housework or farm or home unpaid family work) at a specific day's work for a specific day's pay or commodity paid to the family income.

Col. 24. DID THIS PERSON HAVE A JOB? Enter "Yes" for a person not seeking work who has a new job, or a professional man who has left his office during week of March 24-30 for any of the following reasons: Vacation, temporary illness, industrial dispute, layoff not exceeding 4 weeks without pay, expectation of return to a specific day's layoff due to temporarily bad weather conditions.

Cols. 26 and 27. CLASS OF WORKER:

Wage, salary, or commission worker in private work	PW
Wage or salary worker in Government work	GW
Employer	E
Working on own account	OA
Unpaid family worker	NF

Col. 41. WAR OR MILITARY SERVICE:

World War I	W
Regular Army, Navy, Marines	R
Spanish-American War	S
Regular establishments, Army, Navy, Marines (Philippine insurrection service only)	P
Other war or expedition	Ot

FIGURE 2.

When this is not possible, the type of instructions given above is desirable.

One of the commonest faults in the wording of schedules is the subjective type of answer an item suggests or allows. It is objectionable, for example, to have items in a housing schedule on heat, light, or ventilation checked with such subjective and relative terms as "good," "fair," "bad." It is preferable for the schedule to specify objectively all the general types of heat such as, steam, hot water, gas, oil, electricity, coal stove, etc., and to check the type which applies. Lighting may likewise be classified as natural and artificial, and it may be further sub-classified as direct or indirect. If more accurate information is desired, the relation of window space to floor space, the number of lamps and the power, etc., may be specified. In this way the investigations of different workers will be strictly comparable and subject to quantitative measurement, which would be far from true of reports in terms of "good," "fair," "bad," etc. *Whenever possible the questions or items should be so formulated as to permit the recording of the answer or the observation with a check mark, a cross, crossing out or underscoring words, a number, or "Yes," or "No."* The sample schedule (Fig. 3) reproduced in this chapter is defective in this respect in several instances. (See last column, top, especially.) *Leading questions and questions requiring computations on the part of the informant should be avoided.* We shall return to this subject in the next chapter.

The unrevised questions in the accompanying illustration (Fig. 4) further illustrate the failure to apply some of the principles set forth above. These questions were selected from a series of schedules on church and community affairs.[4] The indefinite, qualitative, and subjective nature of the answers which they call for, or at least permit, would make the results of an investigation conducted with such a schedule largely worthless.

The revision which follows each original question was formulated by a group of students as a laboratory exercise in the application of some of the principles concerning the

[4] F. H. F. Ritchie, *How to Study Your Association and the Community*, (Association Press, New York, 1926).

DEPARTMENT OF LABOR AND INDUSTRIES
BUREAU OF WOMEN AND CHILDREN.
EMPLOYMENT OF WOMEN AND SANITATION.

Firm ____ Address ____ Business ____

Interviewed ____ Position ____

Total No. employes ____ Kind of Bldg ____ Size in stories ____
Total adult male ____ Floors occupied ____
Total adult female ____ Factory laws posted ____
Total children—male ____ No. stairways ____
Total children—female ____ Stairways screened ____
Regular Sunday workers ____ Enclosed ____ Railed ____
Regular night workers ____ No. of fire escapes ____
Employes working seven-day week ____ Kind of fire escapes ____
No. piece workers ____ Metal door ____ Wire glass ____
No. time workers ____ Exits obstructed ____ No. ____
Home Work: No. ____ Kind ____ Aisles obstructed ____
Record of addresses ____ Doors open out ____
Record satisfactory ____ Elevators, No. ____ Signal used ____
How protected ____
Dangerous material used in m'f'g ____ Chemical extinguishers ____
When recharged ____
Waste disposal: safe ____ Insanitary ____ Stand pipes ____
Work dangerous ____ Dimensions ____
Strain ____ 400 cu. ft. per employe ____
Female or children oiling machines ____
Female or children cleaning machines ____

Work room clean ____
Work room light ____
Light largely artificial ____
Kind of artificial light ____
Ventilation good ____
Ventilation artificial ____
Heat ____ Dust ____
Excessive moisture ____
Basement sanitary ____
Seats: Provided? ____ Use allowed ____
Dressing Room: Provided? ____ Needed ____ Clean ____
Lunch room ____
Rest Room ____
First aid ____
Toilets: Separate ____ Sanitary ____ Clean ____ No. ____
Individual Towels ____

Date ____ No. ____ No. of orders issued ____ Inspector ____

Time.
Office record kept ____ Busy days ____ Special inspection needed for:
Office record reliable ____ Machinery ____
Schedule posted ____
Overtime ____ Busy months ____ Fire escape ____
Night work ____
Night work hours ____ Exemptions apply ____ Ventilation ____
Sunday work ____ Exemptions apply how ____
Sunday work hours ____

From A. M. to P. M.	Lunch time.	Extra time off.	Extra time on.	Total day	Total week.	Wages
Sun.						
Mon.						
Tues.						
Wed.						
Thur.						
Fri.						Supplementary Wage
Sat.						Report ____

FIGURE 3. Schedule for the Investigation of an Industrial Establishment.

construction of schedules as discussed above. While these revisions are subject to further improvement, and while their form in many cases would be governed to some extent by the particular church, they represent a vast improvement over the original questions.

4. *Physical Aspects of Schedule*[5]

(*a*) *Size and shape.* — Usually it is desirable to adopt a size and shape which fits some standard filing cabinet, *e.g.*, 5 x 8 or 8½ x 11. A size which does not require folding is generally preferable, when the purpose can be served

[5] For further discussion of this subject as regards mail questionnaires, see Chap. VII.

by a relatively brief schedule. There are cases where larger sheets are used to good advantage, however. For example, the Federal Census of Population (1940) employs sheets 18½ x 24, with items on both sides. Each sheet has room for the recording of the data on 80 persons. The size of the schedule will be determined chiefly by the number of items it contains and the uses or methods of handling planned for the schedules when filled out.

(*b*) *Material and color.* — The grade or kind of paper to be employed should be governed by the degree of permanency and the amount of handling expected for the schedules when filled out. In any case, the paper should take ink easily and be of such grade as not to tear easily or cause other trouble to the enumerator. For some purposes cards are very desirable. They are easy to write on when held in the hands of the enumerator, and if the schedules are planned for ready reference in a card file, their rigidity greatly facilitates locating, handling, sorting. On the other hand, they are more expensive and occupy more space than thinner and more flexible paper. Color is of importance chiefly as a classification device.

(*c*) *Arrangement of items.* — The arrangement of the items on the schedule is of importance from four points of view: (1) The items should be so grouped and spaced as to facilitate the ease with which the eye follows a line or column, or is able to locate the part and subject-matter desired. Titles, headings, and sub-headings are important in this connection. The relative size of the types and the position of the headings should indicate the relationships of different parts of the schedule. (2) The items should be so grouped that related items, or items generally called for together when the schedules are filed or tabulated, will occur in the same part of the schedule. Such arrangement may be a great convenience, if the schedules are used for reference purposes, and may save much time in editing and tabulation, if they are to be handled statistically. (3) Each item should be numbered or lettered in a distinctive way so that it may readily be identified and compared with the same item in other schedules. Each schedule likewise should bear sufficient identification marks to permit its replacement in the proper place if it should be filed out of order or become separated from the filing place of the other schedules. This identification may be in the form of numbers, names, dates, or titles. (4) Undue crowding of items should be avoided.

FIGURE 4. — EXAMPLES OF DESIRABLE REVISIONS OF QUESTIONS APPEARING IN A QUESTIONNAIRE ON CHURCH AND COMMUNITY AFFAIRS

I. Do you go to church regularly?.....

Revision:

(*a*) Do you go to church at all? No..... Yes.....
(*b*) On which of the following special holidays have you attended church within the past year:

1. New Year's.... 2. Good Friday.... 3. Easter....
4. Christmas..... 5. Others.....

(*c*) Indicate your attendance at the following:

Number of Times Attended	*Past Month*	*Past Year*
1. Sunday morning service		
2. Sunday evening service		
3. Mid-Week service		
4. Sabbath School		
5. Social Meetings		
6. Others		

II. Do you contribute to any worthy causes?

Revision:

Indicate the approximate amount of your contribution to the following causes:

Causes	*Per Week*	*Per Month*	*Per Year*
1. Your Church			
2. Missions			
3. Community Chest			
4. Y. M. C. A.			
5. Y. W. C. A.			
6. Salvation Army			
7. Red Cross			
8. Hospitals			
9. Children's Homes			
10. Homes for the Aged			
11. Beggars			
12. Others (specify)			

III. Are you active in church work?.....

Revision:

Check the church activities you have been engaged in during the past year:

(*a*) Sabbath School teacher.....
(*b*) Officer of the Sabbath School.....
(*c*) Church official.....
  1. Elder.....
  2. Trustee.....
  3. Deacon.....
  4. Others.....
(*d*) Choir.....
(*e*) Officer of or worker in other church activities
  1. Boys' club work.....
  2. Girls' club work.....
  3. Church socials.....
  4. Others.....

IV. To what extent has this series of meetings been interesting?

Revision:

1. How many of this series of meetings did you attend? .....
2. Are you in favor of another series of such meetings? .....

V. How happy in their work do the men seem to be?

Revision:

1. Total number of men employed.....
2. Total number of men employed
  (*a*) 5 years or more.....
  (*b*) 3 to 5 years.....
  (*c*) 1 to 3 years.....
  (*d*) Less than 1 year.....
  (*e*) Less than 3 months.....
3. Number of complaints within past year of
  (*a*) Working hours .....
  (*b*) Wages .....
  (*c*) Other conditions.....

VI. To what extent has newspaper publicity been used?

Revision:

1. Total number of square inches of newspaper publicity received last year.....

	Sq. inches	Annual cost	Total circulation
(*a*) Paid Advertisement	.....	.....	.....
(*b*) News notes	.....	.....	.....
(*c*) Other free publicity	.....	.....	.....

VII. Is the room well lighted?

Revision:

1. *Natural light:*
   (*a*) Total sq. ft. window space.....
   (*b*) Total sq. ft. skylight.....
   (*c*) Total sq. ft. floor space.....
   (*d*) Total sq. ft. window space, opening on other rooms or on porch.....
   (*e*) *Height* (ft.) *Distance* (ft.) of next building.

North	.....	.....
South	.....	.....
East	.....	.....
West	.....	.....

2. *Artificial light:*
   (*a*) No. of oil lamps.....
   (*b*) No. of gas lamps.....
   (*c*) No. of electric lamps..... (*d*) Total wattage.....

In summary, we may say that the construction of a schedule involves attention to four distinct steps: (1) A list of questions to which answers are sought; (2) A set of dummy tables which, if supplied with the proper data, would answer these questions; (3) The arrangement of the questions or items in the schedule, with attention to physical appearance, wording, logical order, etc.; (4) Instructions to enumerators.

The most frequently neglected of these steps is the second, namely, adequate foresight as to exactly how the questions included in the schedule are to be tabulated so as to yield the answers sought. The subject of tabulation is a large and important one, and must be left to textbooks on statistics.[6] It is a subject intimately related, however, to the construction of schedules, and attention should therefore be called to this fact in the present connection. If standard Hollerith or I.B.M. electric card-sorting and counting machines are to be used, as in large-scale studies involving thousands of schedules and many interrelations, code-symbols for each possible type of answer should be inserted in the schedule, thus greatly facilitating the editing and card-punching necessary for machine tabulation. The standard machine punch card

[6] A good brief treatise including illustrations of machine methods, will be found in F. E. Croxton and D. J. Cowden, *Applied General Statistics*, (Prentice-Hall, 1940), ch. 2. See also, Katherine Phelps, "A Flexible Method of Hand Tabulation," *Jour. Marketing*, Jan., 1939, p. 265.

contains eighty columns and twelve rows. Each possible answer is assigned a code-symbol corresponding to a particular column and row on the card. Each card is then punched to correspond to the answers on a particular schedule.

F. THE TESTING OF SCHEDULES

Even with the most careful preliminary preparation and criticism, most schedules have defects which only actual use in the field will reveal. For this reason it is always advisable to try out a new schedule in the field, and to revise it on the basis of the discovered defects, before going ahead with it on the whole investigation. This will usually result in the correction of errors which would have impaired seriously the value of the study. We shall discuss this subject more fully in the next chapter.

G. THE FILLING OUT OF SCHEDULES

The purpose of many of the principles of schedule-making which have been discussed above is to insure that they will be intelligibly filled out. One of the principal points we have emphasized is that each item should be unmistakable as to what information is wanted and in what form the answer is desired. In short, one of the chief considerations in constructing a schedule is to make it as nearly fool-proof as possible. But the greatest possible care in the wording of the schedule will not eliminate the need for care in filling it out. Unfortunately, no schedule has yet been devised which will automatically record the facts we want observed and recorded. It requires intelligent treatment on the part of the investigator or the informant, as the case may be. It is an unfortunate fact that in addition to the inadequacy of our public and quasi-public records from the standpoint of structure and completeness, the practical value of these records is further vitiated by ignorance and carelessness in filling them out.

In many cases, carelessness in filling out schedules is due to the fact that the person charged with this duty has no knowledge of the significance of the data he is asked to provide or of what possible usefulness they may be when compiled. As a result, such persons are usually indifferent, and

frequently hostile, to what appears to them a futile task. One frequently finds this attitude among public officials, teachers, and social workers. To the active worker, record-keeping is often "uninteresting," and since each record becomes merged with thousands of others, and perhaps is summarized in a cold average of some kind, many officials and workers are bored with the task of filling out schedules. With this attitude prevalent, it is not surprising that even in states where the law provides fairly adequate schedules for recording vital data, there is the most shocking indifference and incompetence evident in record-keeping. The general defense is that no one except professors, researchers, and other snoopers ever read the records anyway, and it is therefore felt that it is useless to take any time or trouble in making them intelligible.

Almost any collection of a large number of schedules that have been filled out by a number of comparatively unskilled people contain a large number of pathetic and humorous examples of ignorance and incompetency. Whipple cites the following illustrations (some of them perhaps intentionally absurd) of the recording of the causes of death sometimes encountered: [7]

"Went to bed feeling well, but woke up dead."
"Died suddenly, nothing serious."
"Deceased had never been fatally sick."
"Last illness caused by chronic rheumatism but was cured before death."
"Deceased died from blood poison, caused by a broken ankle, which is remarkable, as the automobile struck him between the lamp and the radiator."
"Typhoid fever, bronchitis, pneumonia and a miscarriage."
"Death caused by five doctors."
"Vital Statistics."

The above illustrations are doubtless extreme cases. Less absurd but no less serious errors may easily be found in most collections of records. The most common defects in filling out schedules may be summarized as follows:

1. *Defects in legibility.*—If a word, letter, numeral or symbol cannot be read at all, or may be mistaken for some other symbol, this destroys the value of the record. While

[7] G. C. Whipple, *Vital Statistics* (Second edition, Wiley, 1923), p. 349.

the wording of the schedule may eliminate as far as possible the need for writing, it cannot guard against illegibility of marks or symbols made by the field worker. Reasonable care alone can prevent this fault.

2. *Inadequacy or incompleteness of the matter filled in.* — Sometimes the investigator, in filling out his schedule, will use abbreviations which are intelligible only to himself, and to himself only temporarily. This may make the record useless in many cases. If other than standard abbreviations must be used, a key should be provided if their meaning is at all doubtful. Where the schedule requires the names of persons for identification or other reasons, confusion is often caused by incompleteness. Designations such as "Dr. Smith," "Mr. Jones," etc., are usually nearly worthless. The first name and an initial, if possible, should always be secured. Care should also be exercised in securing the exact spelling of names and addresses. When dates are called for on the schedule, care in securing accuracy should be observed. They may become the basis of legal, financial, or other considerations of the greatest importance.

3. *Check marks and empty spaces.* — In many cases, a check mark in a space of a schedule may have at least three meanings: "Yes," "No," or "Unknown." Usually this ambiguity can be avoided by the wording of the schedule, or by specific instructions. Thus a schedule which inquires into marital status in this way: "Married? . . ." will perhaps ordinarily get an answer of "no" or "yes." But frequently it will be answered by a check mark to indicate that the informant is married, or by a dash to indicate that he is not. Where single words are used as questions in this way, the safer way is to enumerate the alternatives and indicate that a check mark is desired: Thus,

Married .....
Single .....
Widowed .....
Divorced .....
Separated .....

This form is also preferable to the form which has a series of words, calling either for the underscoring of the one it

is desired to indicate, or for the crossing out of all except the one it is desired to indicate. Lack of uniformity in designation, which results from the presence of such alternatives, causes confusion and misinterpretation. Spaces that cannot be filled for lack of information, or because they become irrelevant on account of a certain answer to a previous question, should not be filled with a check mark or other symbols, for they may lead to misinterpretation. A vacant space should be filled with a horizontal line to indicate that the space was not left vacant through oversight, and also to prevent subsequent insertions of data by someone else for purposes of corrupting the report. It also assists the informant or the investigator in looking over his schedule for omissions, and gives assurance to the editors or users of the report that the omission was due to absence of information and not to oversight. The absence of information in a schedule may raise questions of good faith either on the part of the informant or on the part of the registrar, and if the space has been filled with a line the defense of oversight cannot be so easily raised. For all of these reasons, therefore, it is generally desirable *to fill all unused spaces with a horizontal line extending the full length of the space.*

The final test of desirable ways of filling out a schedule should be: *Do the marks convey the desired information unmistakably?* Through a little care and the observation of the simple rules stated above, the accuracy and usefulness of information contained on schedules could be greatly increased.

H. ADVANTAGES OF PERSONAL INVESTIGATION

All of the types of schedule discussed in this chapter assume, first, that only the more objective factual data are sought. Secondly, it is assumed that these schedules will be filled out (*a*) by a more or less skilled field worker who either records the data from direct observation or gets it first-hand from interviews with other people; or (*b*) by the informant without personal contact with the investigator. As has already been noted, the former type of investigation does not require that a schedule should be as detailed in its instructions for filling out, nor indeed as fool-proof generally, as the

questionnaire type which is sent out or distributed for individuals themselves to fill out. When the extent of the inquiry, the time, and cost permit, the personal investigation is usually preferable and has many advantages. First, it permits the use of more complex and more numerous questions. When the schedule is filled out by a skilled investigator complex questions can be explained in whatever terms are necessary to elicit the true answer. Several related and corroboratory questions may be asked orally to secure a single answer. Second, a field investigator can be trained so as to avoid the errors to which the inexperienced or untrained person is subject. Field workers can also be selected on the basis of special advantages of personality and address which will secure responses from individuals who could not be reached at all by less personal approaches, such as letters. These special qualifications of field workers will be considered in a later chapter. Finally, more complete returns can usually be secured through personal investigation.

The advantages of personal investigation are perhaps present in the highest degree where the entire study is performed by a single investigator, provided he is animated by scientific ideals. This usually insures uniformity of approach and interpretation throughout the study as well as the steadying influence of sole responsibility for the accuracy of the results. In many cases it may also mean that the study profits from the enthusiasm and care of a person deeply interested in the results. Of course, there may also be danger in this personal interest. A person may become so enamored of his hypothesis as to render him incapable of impartial and objective observation. But even this is a failing which is least likely to be present in a person of scientific training and ideals. Most field investigations will, of course, require a number of interviewers.

A great many of the data of sociology, furthermore, have to be secured at present through less ideal methods than personal investigation by a trained scientist in the field. In such cases, reliable results may still be secured if care is exercised in impersonal approaches. Many important social data are of a less concrete type than those contemplated in the present chapter. We turn our attention in a subsequent chapter to

methods of securing and interpreting the less tangible social data, such as opinions and attitudes.

I. CONCLUSION

We have considered in this chapter the schedule in general as an instrument for aiding the senses in social observations. Our attention has been devoted chiefly to the types of schedules which assume (*a*) that a skilled investigator will use them, mainly as convenient aids to his memory, and (*b*) that the data called for are of a relatively objective, factual sort. Sometimes, however, we wish to prepare schedules which, in addition to the above purposes, will serve as a set of stimuli to elicit the cooperation of the person from whom information is sought, and as a reliable index to that person's feelings, attitudes and probable behavior in stated circumstances. These purposes are also sometimes served, of course, by the types of schedules discussed in the present chapter. More frequently the questionnaire designed for impersonal distribution is used in the latter type of inquiry. We turn, therefore, to a consideration of the questionnaire as a special kind of schedule.

J. SUGGESTIONS FOR FURTHER STUDY

The best supplementary material for this chapter is the examination of large numbers of schedules and their appraisal from the point of view of the principles discussed above. For this purpose the instructor will do well to collect examples of the blanks, forms, registration cards, etc., used by local, state, and national governments. In addition to the various schedules used by the Bureau of the Census, the Bureau of Labor Statistics, and the Department of Health will provide examples of excellent schedules. As a matter of contrast, a collection of the schedules used in local agencies will usually provide plenty of material for discussion and improvement.

Good examples of schedules that have been developed for field use in special studies will be found in the publications of the Works Progress Administration, Division of Social Re-

search. See, for example, *Urban Workers on Relief,* Research Monograph IV (1936), p. 104, Part I; *Part-Time Farming in the Southeast,* Research Monograph IX, 1937, Appendix E; *Rural Migration in the United States,* Research Monograph XIX, 1939, pp. 170-71; *Migrant Families,* Research Monograph XVIII, 1938, Appendix C.

A very good chapter on the general subject of schedule-making will be found in Lyndon O. Brown, *Market Research and Analysis,* (Ronald Press, 1937).

Other types of schedules will be referred to in the next two chapters.

CHAPTER VII

THE QUESTIONNAIRE

Have you ever noticed this: That people never answer what you say? They answer what you mean, or what they think you mean. Suppose one lady says to another in a country house: 'Is anybody staying with you?' The lady does not answer: 'Yes the butler, the three footmen, the parlor maid,' and so on, though the parlor maid may be in the room, or the butler behind her chair. She says: 'There is nobody staying with us,' meaning nobody of the sort you mean. But suppose a doctor inquiring into an epidemic asks, 'Who is staying in the house?' then the lady will remember the butler, the parlor maid, and the rest. All language is used like that; *you never get a question answered literally, even when you get it answered truly.* — *G. K. Chesterton.*[1]

A. THE QUESTIONNAIRE AS A SPECIAL TYPE OF SCHEDULE

The principles that were set forth in the preceding chapter regarding schedules are also in general applicable to the questionnaire. There are, however, a number of additional considerations which should be kept in mind in the construction of questionnaires, especially when they are to be filled out without the direct personal solicitation and aid of an investigator. Many data are at present collected by this method in the social sciences, to say nothing of the information secured in this way by governmental and private agencies and business concerns.

Among the reasons for its extensive use is the possibility of carrying out this kind of inquiry with a small staff and consequently, if the results are satisfactory, at a low cost. The latter advantage varies considerably in different types of inquiry. In any case, cost can be appraised only in relation to the comparative adequacy and reliability of the results secured.[2] Another consideration in favor of the mail questionnaire is its impersonal nature. It avoids the variety of possibly irrelevant stimuli involved in personal contact. It also preserves the anonymity of the informant in inquiries where this is desirable and may be an important condition

[1] *The Invisible Man.* Quoted in P. F. Lazarsfeld, "The Art of Asking *Why*. Three Principles Underlying the Formulation of Questionnaires," *Nat. Marketing Rev.*, 1:5, 1935.

[2] The advantage of low cost has been denied by some authorities. See footnote 54, p. 208.

for securing his cooperation. As one of the more common instruments of social inquiry, the principles governing the construction and testing of questionnaires and the interpretation of results secured from their use, therefore, is a subject of considerable importance.

Fundamentally, the questionnaire is a set of stimuli to which literate people are exposed in order to observe their verbal behavior under these stimuli. Incidentally we may also observe some of the other behavior of the informants, such as the frequency with which they will return the questionnaire under stated conditions, and how this behavior and, of course, their replies, correlate with other characteristics of the respondents. As to subject matter, questionnaires may be divided for our purpose into two large categories: (1) Questionnaires of fact, which merely ask for information subject to corroboration regardless of the informant's opinions or attitudes, *e.g.,* income tax blanks. (2) Questionnaires of opinion, and attitude, in which the informant's preferences are the basic data sought. We shall be concerned in this chapter with both of these kinds of schedules with regard to principles of construction, distribution, and securing the returns. We shall also assume for the most part that the questionnaire is to be answered by mail. Some special considerations governing the attitude questionnaire and the more elaborate tests and scales are reserved for the next chapter.

B. THE SOLICITATION OF COOPERATION

The request to fill out a questionnaire usually should be in the form of a personal letter in order to take advantage of the already established habit of answering this kind of mail. The writing of persuasive letters is an art that has received much attention from sales organizations and others. The same general principles apply to letters soliciting information by questionnaire. The first thing that attracts attention is the stationery and the letter-head. These features of a letter correspond perhaps to personal appearance in an interview. The recipient of a letter draws certain immediate conclusions as to the character, respectability, and importance of the investigation on the basis of this first impression.

The letter-head also provides a desirable way of saying at the outset some things that the recipient very much wants to know even before reading the letter, and which will influence the attitude with which he reads it. These are the questions of "Who is approaching me?" and "For what purpose?" The letter-head of a university, research or civic organization with prestige in the area to be investigated, perhaps with a subtitle indicating the purpose of the inquiry, together with the names of prominent individuals sponsoring the project, answers these important questions. The stationery and the letter-head, therefore, may determine at the outset whether the recipient gives further attention to the letter or throws it and the questionnaire away.[3]

The nature of the appeal to be made in the letter is the next problem. The recipient must be motivated and provided with an incentive that will overcome the reluctance most people feel toward writing according to rules, not to mention the fact that filling a questionnaire takes time which might otherwise be spent on pleasanter activities. Sletto experimented, in a carefully controlled study, with three types of letters requesting 1600 former university students to answer a long questionnaire. His procedure and results are of interest:

"The first letter called upon the individual to help improve education for the thousands of young people who would be entering the University during the year. The second letter directed attention to the changes occurring in education and requested help to guide these changes 'in the right direction.' The third letter challenged the recipient to help do something that people say 'can't be done.' The letter explained that many people believed a study like this could not succeed because former university students are too much concerned with their personal affairs or too uninterested in education to respond to a lengthy questionnaire. . . The first letter, containing the altruistic appeal to help improve education for the thousands of young people entering the University, yielded the highest returns, amounting to 67 per-

[3] R. Watson ("Investigations by Mail," *Market Research*, 7:11-17, 1937), found in an experimental study no significant differences in number of returns from printed as compared with multigraphed letter-heads, but found that special letter-heads rather than a mere address was helpful.

cent. The second letter yielded 64 percent, and the third letter 60 percent. The first letter was, therefore, adopted for use when the final questionnaire was sent to the final group. The differences in the proportion of returns yielded by the three letters are not large enough to yield critical ratios that are statistically significant according to conventional standards. The superiority of the first letter in producing returns was, nevertheless, consistently maintained throughout the time period of the returns. Since the pretesting of questionnaires will normally involve small numbers of cases, we must expect most 'true' differences to yield small critical ratios."[4]

These results are corroborated by market research experience in which letters merely appealing for information as a personal favor have been more productive than letters offering a trivial reward or stressing some alleged advantage to the respondent for answering.[5] Generally speaking, we may say that the chief resistance to be overcome in mail questionnaires as in personal interviews is a suspicion that someone is trying to sell something. It is well, therefore, to make clear at the outset by all possible means that *information* only is sought, and that no tricks of salesmanship are being employed.

Among the important inducements to reply is a stamped and addressed return envelope. The convenience of this arrangement, its favorable psychological effect, the amount of the postage, plus perhaps a reluctance to throw away a stamped envelope, appear to be among the factors involved. The printed "Business Reply Envelope," on which the postage is paid by the person or organization to whom the envelope is addressed, has been extensively used in recent years. Since postage is paid on these envelopes only if they are returned, they are assumed to represent a considerable economy for organizations sending out many return envelopes (although the Business Reply Envelope requires 4 cents). Watson found, however, in one market investigation that inquiries accompanied by the old-type stamped envelope se-

[4] R. F. Sletto, "Pretesting of Questionnaires," *Amer. Sociol. Rev.*, 5:193-200, 1940.

[5] R. Watson, *op. cit.*

cured twice as many replies as the Business Reply Envelope.[6] This investigator remarks: "Women apparently hated to see that 3-cent stamp go to waste." If the findings of this study are generally valid, the economy of the business reply envelope is doubtful. The unmistakable results in the study mentioned illustrate also the unsuspected psychological factors that may determine whether a questionnaire is returned or not.

One difficulty with the formulation of an effective appeal to accompany questionnaires is that these requests must be formulated so as to have *general* appeal. In an interview the approach can be modified to suit the individual case. If the questionnaire is to be circulated in a selected group, a knowledge of the interests, prejudices, and intelligence of that group will be of great help in formulating an effective appeal. Where the group is heterogeneous in character, general knowledge of human nature and the experience of other similar investigations must be used as a guide. As in the formulation of the questionnaire, the least intelligent, rather than the average informant, must be kept in mind.

C. THE CONTENT AND STRUCTURE OF THE QUESTIONNAIRE

All of our remarks in the preceding chapter regarding (1) the importance in schedules of having clearly in mind the specific questions to be answered, (2) clarity of wording, and (3) discretion in the choice and arrangement of items, apply with even greater emphasis to questionnaires. The questions of explanation and interpretation that almost invariably arise in an interview must be anticipated and answered by the questionnaire itself, preferably by a wording and arrangement of items so unmistakable as to prevent the questions from arising. When this is impossible, detailed and unmistakable instructions must accompany the questionnaire. If possible, such instructions as are necessary should be inserted at the place needed rather than at the beginning or the end of the questionnaire.

When we have decided upon the specific purposes of the inquiry, the first task is to formulate questions calculated to yield the desired information. These should be as few as is consistent with the ends sought. Beyond this consideration,

[6] *Ibid.*

A MESSAGE FROM THE PRESIDENT OF THE UNITED STATES

THE WHITE HOUSE
WASHINGTON

TO EVERY WORKER:

If you are unemployed or partly unemployed and are able to work and are seeking work, please fill out this report card *right away* and mail it before midnight, Saturday, November 20, 1937. No postage stamp is needed.

The Congress directed me to take this census. It is important to the unemployed and to everyone in this land that the census be complete, honest, and accurate. If you give me the facts I shall try to use them for the benefit of all who need and want work and do not now have it.

Franklin D. Roosevelt

(SEE INSIDE)

UNEMPLOYMENT REPORT CARD
Please fill out this card if you are totally or partly unemployed
Answer all questions promptly

It is important that EVERY totally or partly unemployed worker in your household fill out a SEPARATE report card. Additional cards can be secured from your postman or post office. You can get help in answering these questions, if you need it, at any post office or from any postal employee.

1. Print full name ..... (First) (Middle) (Last) — Do you live on a farm? ..... (Yes or no)
   Print full address ..... (Street and number or R. F. D.) (City, town, or village) (County) (State)
2. Are you: (Check only one)
   (*a*) Totally unemployed and want work? ☐
   (*b*) Partly employed and want **more** work? ☐
   (*c*) Working at WPA, NYA, CCC, or other emergency work? ☐
3. Are you able to work? (Check one) Yes ☐ No ☐
4. Age at last birthday ..... ____ years
5. Color or race (Check one) White ☐ Negro ☐ Other ☐
6. Sex (Check) Male ☐ Female ☐
7. How many hours did you work last week? ____ hours (If none, write "None")
   Enter total number of hours worked at ALL jobs (except WPA, NYA, CCC, or other emergency work) during the week from Sunday, November 7, through Saturday, November 13.
8. How many weeks did you work in the last 12 months? ____ weeks (If none, write "None")
   Enter total number of weeks worked at ALL jobs (except WPA, NYA, CCC, or other emergency work) during the 12 months from November 1, 1936, through October 1937.
9. What is your occupation, or kind of work? ..... (Examples: Tool maker, waitress, watchman, farm laborer, etc.)
   The occupation which you give in answer to this question (number 9) describes the work you do. The answer to the next question (number 10) describes the kind of business or industry in which the work is done. Enter "New worker" in both questions 9 and 10 if you have never had a steady job, and want work. Do not give the name of your company or employer as an answer to question 10.
10. Kind of business or industry in which you did, or are doing, this kind of work? ..... (Examples: Machine shop, restaurant, railroad, cotton farm, etc.)
11. How many **other** workers are there in **your family** living in the same household with you? ..... ____ workers
    Count all persons working for pay or profit, or wanting work, **except yourself.**
12. How many of these workers are: ..... (*a*) Totally unemployed and want work? ..... ____ workers
    (*b*) Partly employed and want more work? ..... ____ workers
    (*c*) Working at WPA, NYA, CCC, or other emergency work? ..... ____ workers
    Do not count yourself

    Note.—Every totally or partly unemployed worker in your family should make out and return a separate report card.
13. How many persons are mainly dependent on you for support? (Do not count yourself) ..... ____ dependents
14. What was your individual total income, cash and other, last week? ..... $____
    Do not include payments from relief, WPA, NYA, CCC, or other emergency work.

Sign here .....

Mail this card before midnight November 20, 1937. No postage stamp is needed [OVER]

FIGURE 5. Example of a Questionnaire.

no rule can be laid down as to maximum, minimum, or optimum numbers of questions. The other variables determining whether or not a questionnaire will be returned are too numerous to enable us to estimate with much reliability the degree to which the length of the schedule was a determining factor. Sletto selected 300 former university students and divided them into three subgroups, every third name from an alphabetical list being assigned to each group. He describes the results as follows:

"The first subgroup was mailed a 10-page questionnaire relating to their vocational activities, interests, and needs. The second subgroup was mailed a 25-page questionnaire relating to their socio-civic activities, interests, and attitudes. The third subgroup was mailed both of these questionnaires totaling 35 pages in length. These mimeographed questionnaires were of the check-list type, and all were mailed on the same day. Utilizing three follow-up requests, the proportion of returns was 68 percent from the group sent the 10 pages and 60 percent from that sent the 25 pages. The group sent both questionnaires in the same envelope yielded 63 percent returns. Since the group sent the 35 pages yielded a larger proportion of returns than the group sent the 25 pages, and only five percent fewer returns than the group sent the 10 pages, it was concluded that the factor of length, between the limits of 10 and 35 pages, was not likely to affect returns to the final questionnaire by more than five to ten percent. There is reason to think that the difference between the returns to the 10 pages and to the 25 pages was less a result of length than of the interest value of the questionnaire content, since the group sent both questionnaires yielded a larger proportion of returns than did the group sent the 25-page questionnaire.

"This first experiment in pretesting suggests that the factor of length is less important than it has generally been assumed to be, insofar as proportion of returns is concerned. However, it may be that the effect of length is more pronounced within the limits of one to ten pages, and that within this range a virtual plateau is reached in questionnaires of the check-list type. If such a situation exists, is there some limit beyond which the returns begin to fall off

sharply when additional pages are added to the questionnaire? Would these same results be gained, if the group were not one consisting of college-trained people? Future research alone can provide us with answers." [7]

Another student of the subject has made the point that it is not so much the length of the questionnaire as the number of unrelated subjects dealt with that is objectionable.[8] If the questionnaire is an integrated whole, and if the questions have obvious relationship to each other and to the avowed purpose of the inquiry, the mere number of questions will seem less important to the person answering. In short, the capacity of the questionnaire to arouse the interest of the person to whom it is presented and to cause him to feel that this is a "sensible" inquiry on a subject of significance, is perhaps of greater importance in determining his willingness to respond than is the length of the questionnaire.

Even if research should reveal that the length of the questionnaire, within limits, is a factor of minor importance in determining the proportion returned, there are, of course, other grounds, such as mere economy of space and labor, for keeping the schedule as brief as is consistent with the results we seek. Apparent exceptions to this rule would be the instances in which it is deemed advisable for psychological reasons to insert a limited number of "waste" questions in order to stimulate interest, facilitate the approach to delicate subjects, or corroborate other questions. For example, a market study of people's preferences in building material as between lumber, marble, concrete, etc., might begin with an inquiry about the person's attitude toward the conservation of natural resources, in the name of which the original approach may have been made, as an appeal to broad social and civic interests. Again, a question about the changing purchasing power of the dollar may serve as a disarming opening in a questionnaire designed really to determine present expenditure habits. An investigation of the use of various methods of contraception began with the question "Would you approve of straight-forward, tactful, inoffensive advertising of

[7] R. F. Sletto, *op. cit.*, pp. 195-96.

[8] R. Franzen, "The Construction of a Questionnaire," *Market Research*, May, 1936, pp. 17-19.

this type of product?" "The question served to draw into a discussion of birth control the very women who were opposed to such discussion, and who would otherwise have been lost from the sampling. Such women, having given a negative response, were then asked how they thought such information should be made available, if not through advertising. In most cases, anxious to defend their stand, they were willing to answer and to proceed with the remainder of the interview."[9] While this case involved replies to questions in an interview rather than in a questionnaire, the principle of the technique is the same.

Even if no "waste" questions, (which need not be tabulated) are deliberately included, it is advisable to begin the questionnaire with inquiries that are unobjectionable, familiar, and easily answered. The routine items of sex, nationality, length of residence in present place, etc., satisfy these requirements. If the first questions appear easy, a person may get started answering and, if so, an important objective has been achieved.

The arrangement of the questions in the body of the schedule is also a matter of some importance. A question which is difficult, objectionable, or sure to arouse antagonism at the beginning of a questionnaire may be answered without any hesitation after an adequate foundation has been laid for it and after the person answering has "warmed up." The reinforcing effect of wishing to finish something that has been started can be counted on to secure answers to questions in the last part of the questionnaire which, if asked at the outset, would probably have caused refusal to return the schedule. The whole psychology of questioning is very intricate. Certain aspects of the technique have been very highly developed by the legal profession. Many other aspects and types of interviewing remain to be studied by research workers.

D. WORDING OF QUESTIONS

The order of the questions is very intimately related to their wording, and *vice versa*. A question that is perfectly

[9] J. W. Riley and Matilda White, "The Use of Various Methods of Contraception," *Amer. Sociol. Rev.*, 5:890-903, 1940, p. 891.

clear if it follows several other questions might be unintelligible if asked by itself. That is, the wording necessarily depends in part upon the order in which the questions are asked. What we said in the preceding chapter regarding the wording of schedules in general also applies with special emphasis to the wording of questionnaires. The illustrations given in Chapter VI will also serve to emphasize some elementary considerations in the present connection. The matter will come up again in the next chapter in connection with the measurement of attitudes. We shall confine ourselves, therefore, at present, to a few special cautions in the wording of questionnaires.

Bowley's general rules on this point constitute a good summary of the ends to be aimed at, although they do not tell us much about how to achieve these results. According to this authority, questions should have the following characteristics:

1. Comparatively few in number
2. Require an answer of a number or of a "yes" or "no"
3. Simple enough to be readily understood
4. Such as will be answered without bias
5. Not unnecessarily inquisitorial
6. As far as possible corroboratory
7. Such as directly and unmistakably cover the point of information desired.[10]

The education and general intellectual level of the public from which answers are desired will determine to a large extent what degree of simplicity (*i.e.,* in vocabulary, sentence structure, etc.) is to be employed. The only final guide on this point, as indeed on most others, is adequate pretesting of the questionnaire on sample populations, and the revision of the schedule on this basis. In the preparation of the preliminary forms for such testing, the advice of persons intimately familiar with the group to be studied is much to be desired. Bowley's rules may be supplemented by several specific observations on problems that always arise.

1. *Categorized versus "free" answers.* Is it desirable to list, as far as can be imagined, all possible answers in order to permit replies exclusively by a check mark or by yes or

[10] A. L. Bowley, *Elements of Statistics,* (Fourth edition, King, London, 1920), pp. 18-25.

no? This is one of the most difficult problems in the formulation of questionnaires. Unless the answers sought are sharply and obviously exclusive, and exhaustive, the listing of a variety of possible responses frequently results in the checking of several answers, so that the result may be difficult to interpret. For example, to inquire regarding marital status by requesting a check mark after one of the categories, "married," "single," "divorced," "divorced and remarried," "separated," "widowed," is likely to yield satisfactory results because the categories are exhaustive and mutually exclusive. Consider, on the other hand, the following:

Where did you first meet your college roommate:
(*a*) At preparatory school? . . .
(*b*) In the same community? . . .
(*c*) In the same neighborhood? . . .
(*d*) Other. . .

Obviously the first three of these might be checked. If an attempt is made to list all possible answers in a case of this kind, an absurdly long list would have to be appended. If this practice is adopted in questions of attitude we may suggest to the person receiving the questionnaire many answers which he probably otherwise would never have thought of. In some cases this may be desirable, *e.g.*, when we are studying how people will respond when exposed to certain suggestions. But in many questionnaire inquiries we are probably interested in the person's own formulation of his response. If we allow him to express this response in his own words, difficulties of interpreting his answers are sure to arise. Conversely, if we limit him to our own list of answers we force his response into our prearranged categories.

Most practical inquiries, in the face of this dilemma, adopt a compromise. All questionnaires are necessarily formulated on the basis of some knowledge of the type of answer probable. Such knowledge may be secured from experimental questionnaires permitting free response to questions, without any attempt to categorize the answers in advance. The more intimately the investigator knows his public, the more effectively he can formulate questions and the categories into which most answers will fall. Hence a careful scrutiny of all available previous attempts to study the subject

under consideration should precede the formulation of a schedule. Adequate pretesting of the questionnaire should yield most of the knowledge needed for a formulation of answer categories which will do a minimum of violence to the spontaneous reaction of the respondent. The easy way in constructing the questionnaire is to ask merely the questions and allow the respondents complete freedom as to the words and categories in which they wish to answer. The classification of these responses after the schedules come in then becomes the problem. Except in cases where the possible answers are mutually exclusive and exhaustive, a compromise is usually adopted. The most likely and frequent answers are listed for check, with a category, "others," concluding the list of possible answers. The problem of classifying the responses received under the omnibus category is then the same as mentioned above. But if the responses in this category are few and not too varied, the proportions of the problem are greatly reduced as compared with the situation when no attempt is made to categorize the answers.

2. *"Leading" questions.* Closely related with the problem of categorizing possible answers is the problem of "leading" questions in general. These are questions so phrased as to suggest a certain answer. The salesman in a store approaches the customer with, "What will you have?" rather than, "Will you have something?" He asks, "How many?" instead of "Do you want more than one?" These may seem trivial illustrations, but the phrasings suggest an answer that the questioner desires. Susceptibility to this type of suggestion naturally varies among different people. Sometimes etiquette itself seems to dictate the propriety of giving people the answers they desire. Thus, Stefansson reports that the Eskimos tend to give encouraging reports about lost explorers because of the obvious desire of searching parties to secure such reports.[11] The Chinese are correspondingly po-

[11] V. Stefansson, *Unsolved Mysteries of the Arctic,* (Macmillan, 1939): "If an investigator said to a primitive Eskimo that surely the Franklin party must have killed a lot of caribou, the answer — irrespective of fact — would be that they certainly did, that they killed an awful lot." (p. 118)

Leading questions may, of course, have value if they are asked deliberately and the results are treated accordingly. For example, if affirmative replies are secured even after strong counter-suggestion by the form of the question, this may be a valuable measure of the firmness of a belief or of the resistance to be overcome. *Cf.* R. Franzen, *op. cit.*

lite.[12] Suggestiveness in response to leading questions may vary considerably with age, education, and different cultures. An example of the same phenomenon is found in the "yes" tendency in public opinion polls. For example, it is better to ask, "Is it *desirable* or *undesirable* to balance the budget," than to mention only *one* of the alternatives.

Leading questions are most likely to be employed, sometimes quite inadvertently, when the investigator is eager to have his results prove rather than disprove some cherished thesis. If an investigator of the causes of application for relief should say to a laborer, "I suppose you were forced to seek relief because of the depression?" or, "Unemployment is inevitable under capitalism, isn't it?" the answers would be likely to show a high percentage of affirmative responses, especially if it is assumed that the questioner may have influence in the granting of relief. Such flagrant cases probably do not occur in any study purporting to yield unbiased results. But the principle involved frequently crops up in subtle ways because of the emotional aura attached to certain words. We shall return to this subject in the next chapter. More refined methods of determining gradations of intensity or firmness of beliefs, as well as the undesirability of hypothetical questions, and questions hedged about with too many reservations, will also be considered in that chapter. In the meantime we may summarize here some of the results of actual research into this important field.

3. *The effect of stereotyped words.* We have already referred (p. 66) to the biasing influence of attaching certain words, such as "Red," "radical," "Fascist," "Communist," to questions regarding policy or practice. Thus Menefee found that certain propositions were accepted when not characterized by these labels but were rejected when these labels were attached.[13] Turner[14] warns against the emotional connota-

[12] Carl Crow, *Four Hundred Million Customers,* (Harper, 1937), pp. 132-33.

[13] S. Menefee, "The Effect of Stereotyped Words on Political Judgments," *Amer. Sociol. Rev.,* 1:614-21, 1936. The biasing influence of the word "Fascism" at present in the United States has also been studied by E. Raskin and S. W. Cook ("A Further Investigation of the Measurement of an Attitude toward Fascism," *Jour. Social Psychol.,* 9:201-06, 1938) with similar results. See also R. Stagner, "The Prestige Value of Different Types of Leadership," *Sociol. and Soc. Research,* 25:403-13, 1941.

[14] R. C. Turner, "A Public Opinion Poll on Educational and Other Issues," *Michigan Public Opinion Surveys,* 1939, p. 49.

tions of *capitalistic, save, guarantee, rich,* and *charity.* The Institute for Propaganda Analysis[15] notes a number of words as examples of "name-calling" rather than designations of objective content. Among them are such words as *alien, demagog, dictator, economic royalist, muckraker.*

The names of political parties[16] and political personages, when associated with certain inquiries, also bias the response. If reliable attitudes on *issues alone,* aside from parties or personages, are desired, the names of the latter should be avoided. Thus the question, "Do you approve of the idea of having Thanksgiving a week earlier this year?" received a less favorable response than when the question was worded, "Do you approve of President Roosevelt's idea of having Thanksgiving a week earlier this year?"[17] Likewise, Studenski showed that the wording "spending" as applied to governmental finance arouses antagonistic responses not stimulated when the question is rephrased so as to omit this word.[18]

4. *Ambiguous words.* The different meanings that words have for different people is a source of constant and subtle errors in the interpretation of questionnaire results. The word "kinds" as in, "What kinds of powder do you use?" is sometimes taken to mean what brands; sometimes the answer will be in terms of form (loose or compact); and sometimes in terms of color.[19] Robinson found that the word "publicly" has different meanings to different people in such a question as, "Is the electric utility in your community privately or publicly owned?"[20] Some responded that it was publicly owned because the public owned the shares.

The extensive interest in public opinion polls has greatly stimulated attention to the wording of questions and the results of *systematic analysis of future questionnaire studies*

[15] *Propaganda Analysis,* 1, No. 1, Nov. 1937.

[16] A. B. Blankenship, "The Choice of Words in Poll Questions," *Sociol. and Soc. Research,* 25:12-18, 1940.

[17] *Ibid.* See also E. Roper, "Wording Questions for the Polls," *Public Opinion Quarterly,* 4:129-30, 1940.

[18] P. Studenski, "How Polls Can Mislead," *Harper's Magazine,* 180:80-83, 1939.

[19] A. B. Blankenship, *op. cit.*

[20] C. S. Robinson, "The New Science of Public Opinion Measurement and its Implications for Business," *Harvard Business School Alumni Bulletin,* 1939, pp. 3-8.

will doubtless make available much reliable knowledge at which we can now only guess from scattered studies of the kind here mentioned. A useful guide to vocabulary is E. L. Thorndike's *A Teacher's Word Book of* 20,000 *Words*.[21] It treats the level of difficulty of different words as indicated by the frequency of use in common sources of literature. Only adequate pretesting of the questionnaire with definite checks will reveal to what extent the vocabulary used means the same to the public which we propose to investigate.

E. PHYSICAL AND ESTHETIC ASPECTS

The considerations noted in the preceding chapter (p. 170) with respect to the physical aspects of schedules in general apply also to questionnaires. We have also noted above certain considerations regarding the stationery and printing of the introductory letter. Sletto's report on several experiments on the effects of physical and esthetic factors in questionnaires is illuminating both as an indication of the kind of items to be considered and as a report on one of the first inquiries of its kind.

"Several pretesting experiments were conducted to gain knowledge applicable to enhancing the attractiveness and interest value of the questionnaire. The knowledge acquired in these experiments was applied to select an esthetically pleasing cover design for the questionnaire, to find a title which would arouse interest, to choose an attractive page format, to determine upon a size and style of type that would make the final printed questionnaire easily readable under poor illumination and by people with less than normal vision, and to find photographs of high interest value and appropriateness to illustrate the various sections of the questionnaire.

"One of these experimental studies of esthetic preferences will be cited as representative of this type of pretesting. The purpose of this experiment was to select a page format that would be esthetically pleasing and to ascertain the extent to which esthetic preferences, as they relate to design of questionnaires, are similar from person to person and for similar groups of persons. The spacing of photograph, page title and text, and the character of the decorative lines used to

[21] Columbia University Press, 1931.

integrate these elements of the page, are factors affecting the esthetic impression produced by a page design.

"Ten different page formats, prepared by an artist who had special qualifications for this work, were submitted to one university class containing 23 students and to another containing 24. Art students were avoided in this experiment on the assumption that they were likely to differ more from a cross-section of the former university population than do students in other university classes. Each person in these two classes was requested to select the page he liked best and the one he liked least among the ten submitted. The same photograph, title, and text were found on each page in order to control these factors. The spacing of these elements and the decorative lines employed were the only variables from page to page.

"Students in this experiment were cautioned not to communicate with each other until after their choices had been recorded on slips of paper and collected. When these choices were tabulated, it was found that two of the ten designs were much superior to the others, since they received 28 of the 47 votes as most preferred and only three of the 47 negative votes as least preferred. The two least liked designs were accorded 32 of the 47 negative votes and none of the positive votes. The remaining 19 positive votes and 12 negative votes were distributed among the other six designs. When these 10 designs were ranked according to the number of positive votes obtained in each group, and these ranks were correlated for the two groups, a rank order coefficient of .74 was found. When the designs were ranked according to the number of negative votes, the rank order correlation coefficient was .70. This result indicates that aesthetic preferences are not highly individualistic and erratic, but rather that a strong consensus of preferences exists within homogeneous social groups." [22]

F. PRETESTING OF QUESTIONNAIRES

The final revision of a questionnaire should be made on the basis of actual test on a public as nearly as possible like the one to be investigated. Not even experienced investi-

[22] R. F. Sletto, *op. cit.*, pp. 198-99.

gators thoroughly familiar with the subject, the field, and the technique of questionnaire construction can foresee all of the difficulties of interpretation and response that the recipients of the schedule will bring to his attention. The inexperienced researcher is likely to be impatient with this preliminary work, which may seem like hair-splitting over the meaning of words, and other details. But patience and care in this preliminary work may make all the difference between success and failure, both in the cooperation of the respondents and in the reliability and validity of the results. Sletto reports that more than six months were spent in the construction of the questionnaire described above, every page being revised at least five times before reaching final form.[23] In a later chapter (Chap. IX) we shall give a more detailed account of the construction and pretesting of a scoring schedule. Although this type of schedule differs in important respects from the questionnaire, many of the principles governing the construction of both types are the same.

The general purpose of pretesting a schedule is to determine its *reliability* and *validity*. This is important whether it is possible to improve the schedule for actual use or not, as this knowledge forms the basis for our estimate of the degree of significance to be attached to the results. We have already touched briefly on methods of estimating the reliability of a sample in our chapter on Sampling (Chap. V, p. 147). We shall return to the subject of statistical reliability and errors again in the next chapter. This subject is, however, a major concern of the field of statistics and the student will need to be familiar with the elements of probability theory for a real appreciation and discriminating use of the various tests of reliability. The present treatment, therefore, aims only to call attention to some of the problems involved.

G. RELIABILITY

A schedule, questionnaire, or other instrument is reliable if it gives the same or highly similar results when reapplied to the same or highly similar phenomena. Two census takers canvassing the same community on the same day with the

[23] *Ibid.*, p. 199.

schedule shown in Fig. 2 should bring in the same information. If they do not bring in the same results, the discrepancy must be due to (1) variation in the interviewers' interpretations of the questions; (2) variation in the recording of the answers; or (3) variation in the informants' answers on the two occasions, such as might occur, for example, from the fact that each interviewer interviewed different members of the family. In the case of a questionnaire, we would call it unreliable if it failed to bring in the same information when repeatedly submitted to the same person, assuming that it called for facts which had not changed between the two occasions. In actual studies involving numbers of cases we consider the questionnaire reliable if the percentage answering a certain way or the mean of the distribution and its standard deviation remain fairly stable when the questionnaire is resubmitted to the same population or to another population we believe, from other evidence, to be highly similar. Thus it was found that questionnaires on regularity of radio listening to certain programs, if repeated within 15 days, gave results varying by a maximum of 8 percent.[24] Another research organization re-interviewed 10,000 people in rural areas and found that the reported *average* daily listening for the whole group differed by only five minutes in the two studies.[25] This is not to say that there may not have been great discrepancies in individual reports. The study merely shows that sampled information on listening habits was reliable in given cases, if taken for a group as a whole.

The reliability of questionnaire responses will vary greatly according to subject-matter, the public questioned, and the time elapsing between questionnaires. Responses for an entire group will usually be more stable (reliable) than individual responses. Bain[26] found that college freshmen changed their replies on nearly one-fourth of 3050 items after a period of two and a half months, although they were urged to

[24] Alberta Curtis, "The Reliability of a Report on Listening Habits," *Jour. Applied Psychol.*, 23:127-30, 1939. See also, "Checking the Checkers," *Advertising and Selling*, Dec. 19, 1935. Also "Dependability of Psychological Brand Barometers," *Jour. Applied Psychol.*, 22:1-7, 1938.

[25] Alberta Curtis, *op. cit.*, p. 129.

[26] Read Bain, "Stability in Questionnaire Response," *Amer. Jour. Sociol.*, 37:445-53, 1931.

answer the questions the same way as nearly as possible. The questions covered factual family data, factual personal data, and subjective personal data. Girls showed uniformly greater stability of response than boys, and the whole group showed greater stability on factual personal data than on other types of response. These results were confirmed in a highly similar experiment by Smith.[27] A somewhat similar study by Cavan[28] on eighth-grade children at an interval of only one week showed changes in response on only about 13 percent of the items. The individual stability of questionnaire replies is probably one which has to be estimated in each case rather than according to any general rule, on account of the large number of variables likely to be operating.

Aside from any fault of the schedule, we know that there will be some variation in the results on account of the operation of "pure chance," *i.e.,* multiplicity of small causes which tend to balance each other in a large number of cases, *e.g.,* mere *mistakes* in speech, reading, marking the schedule, etc. The *probability* that an observed difference in the obtained results is due to these chance "accidents," however, can be estimated from the known laws of probability that have been developed from observations of how chance phenomena distribute themselves in frequency, as, for example, the distribution of heads and tails when coins are tossed a large number of times. On this basis we can estimate what the probability is that the observed difference in the results of two or more trials of a questionnaire is due to chance. If the difference secured is so small that the likelihood is high that it would tend to disappear as the number of questionnaires increased, we attribute the discrepancy to "sampling error," measured by the formula given on p. 147. If, however, the discrepancy between the results of two trials of the questionnaire is so large that, by the laws of probability, the likelihood is small that such a difference could be due to mere chance variation, we must look for the reason for the variation in some ambiguity in the questionnaire. In the case of schedules filled in by an interviewer, there is, of course, the addi-

[27] M. Smith, "A Note on Stability in Questionnaire Response," *Amer. Jour. Sociol.,* 38:713-20, 1933.

[28] Ruth S. Cavan, "The Questionnaire in a Sociological Research Project," *Amer. Jour. Sociol.,* 38:721-27, 1933.

tional possibility of *consistent* and *persistent* misinterpretation of a question or a persistent misrecording of the answers by one or several interviewers. By giving a questionnaire to a trial group on different occasions and comparing the results, the less reliable questions, *i.e.,* those showing a lack of *stability* in their responses, can be identified and modified. Corroboratory questions, *e.g.,* asking both "Age" and "Date of Birth," may also be used to a certain extent as a check on reliability.

H. VALIDITY

The *validity* of the results of a questionnaire refers to the extent to which these results correspond to some criterion which we regard as sufficiently valid to accept the agreement of the questionnaire results with it as a test of the validity of the questionnaire results. In some cases such criteria of a final sort are provided by subsequent events. That is, when the election results are predicted by a questionnaire, the questionnaire is validated by the results of the election. More frequently the results of questionnaires are considered valid on the basis of the dangerous criterion that the results correspond with "commonsense" or with what seems "reasonable." In the absence of other criteria, and always in addition to them, commonsense is entitled to consideration. But in view of the fact that systematic questionnaire studies are undertaken because we do not trust the commonsense estimates, we can hardly use the latter as sole criteria of the validity of the former. More justifiably the judgment of experts is relied on. Again, the results of the questionnaire may be compared with results on the same questions obtained by other methods thought likely to yield more valid responses than the questionnaire, such as oral interviews. When the questionnaire asks for certain types of facts, individual check is sometimes possible from reliable public records such as birth certificates, real estate titles, etc.

It is probably true that the estimated validity of questionnaire results is frequently presumptive and inferential rather than direct. We assume high validity when we have guarded against the known possibilities of error, when there is no known reason why informants should intentionally

mislead, and when the results are not inconsistent with other known facts about the respondents or with other measures, the validity of which is accepted. We shall return to the subject of validity in a later chapter in connection with measuring scales. In the end, the validity of all instruments will be determined by their usefulness in serving the discriminatory purposes for which they were invented. These purposes are likely sooner or later to involve practical adjustments of a kind which yield a positive (*i.e.*, generally accepted) answer to the question of validity of our instruments.[29]

I. SENDING OUT THE QUESTIONNAIRE

When all the above precautions have been taken with respect to the letter of solicitation, the content, structure, wording and form of the questionnaire, and when it has been pretested and revised on the basis of revealed weaknesses, we are ready to send it out to the population selected for the inquiry. That population will usually be a sample, and if so, all of the principles of valid sampling discussed in the preceding chapter apply at this point.

It is usually desirable to send out all of the questionnaires at once, or so that they will be received about the same time. Otherwise some recipient may become suspicious that he has been singled out for special inquiry. It also makes it easier to conduct the follow-up work and tends to get the returns in at approximately the same time. As to the time of the week when a questionnaire should arrive in order to stand the best chance of being filled out, Toops advances the theory, based on some data, that questionnaires tend to be filled out during week-ends. "Accordingly, the questionnaire should not arrive at the recipient's home so early in the week that he will lay it aside and forget it, or throw it in the waste-basket for lack of time to answer it immediately, but

[29] For example, the validity of the thermometer is considered established if we find that, when it falls below a certain point, disastrous consequences on plumbing, automobile radiators, etc., always ensue. The theory of the expansion and contraction of matter with changes in temperature is merely the verbal rationale and generalization of the practical experience noted. For further discussion of the validity of measuring instruments, see my *Foundations of Sociology*, (Macmillan, 1939), pp. 446-54.

rather should arrive about Friday night or Saturday morning, whereupon it may receive immediate attention." [30]

J. FOLLOW-UP LETTERS

The time to send follow-up requests will usually be determined by the rate at which returns are coming in. When returns show a distinct falling off, some kind of a follow-up should be brought to bear upon those who have not yet replied. This may be in the form of a letter together with the original schedule, a double postcard [31] with only the more important questions, a telephone call, or an interview. If the telephone or interviews are used, it may be sufficient to take only a random sample of the non-respondents. This involves all the problems of valid sampling, however, and, in the case of telephone interviews, the biasing influence that telephones are associated with economic status.

How many follow-up efforts are desirable is, of course, a matter of how important a larger percentage of replies is considered from the standpoint of the representative nature of the sample secured. This can be determined by a tabulation of the results of some of the replies already received. "One should compute daily the cumulative averages or percentages on the four or five most significant and essential questions of the questionnaire. When these significant averages and percentages vary but little from week to week, one has some reasonable assurance that the follow-up has proceeded about far enough." [32] Toops received replies from 16 out of a remaining 24 as a result of his fifth follow-up letter to former college students.[33] Six follow-up letters secured 100 percent returns on 110 questionnaires to college professors and administrators.[34]

[30] H. A. Toops, "Validating the Questionnaire Method," *Jour. Personnel Research*, 2:157-58, 1923.

[31] These are regular postcards, clipped together, one to be torn off and returned. Sletto (*op. cit.*) found that postcards yield the same frequency of response as letters when used as follow-up stimuli.

[32] Herbert A. Toops, "Validating the Questionnaire Method," *op. cit.*, pp. 153-69. This article also contains suggestive copies of types of follow-up letters.

[33] *Ibid.*, p. 156.

[34] H. A. Toops, "The Returns from Follow-Up Letters to Questionnaires," *Jour. Applied Psychol.*, 10:92-101, 1926.

Stanton[35] secured an increase of 75 percent (from 28.3 percent to 50.2 percent) through his first follow-up to a questionnaire to teachers on the use of radio equipment in the classroom. The same author also reports a 94 percent return on a nine-page discussion and questionnaire as a result of pretesting the questionnaire and three full-length follow-ups[36] (one when 54 percent of the replies had been received, the second when 72 percent, and the third when 89 percent had been returned).

Suchman and McCandless[37] report the results of the following technique: A random list of 600 women was selected from the telephone directory of Cedar Rapids and Iowa City. A questionnaire was sent to each woman. When not more than one return per day was received for five successive days, a second wave of the same questionnaire was sent to all who had not answered. After a similar period of waiting, a random sample of every second person of those who had not answered was called on the telephone for their answer. This sample was considered representative of all those who had not returned the questionnaires.

The results are indicated in Table II.

TABLE II[38]

Proportion of Replies to Successive Waves of a Questionnaire on Radio Programs

Wave	*Number sent*	*Percent replies*
First wave (mail)	600	16.8
Second wave (mail)	490	34.1
Residual sample (telephone)	141	97.2

The same authors report the following experience with a random group of 900 subscribers (selected from a list of 10,000) to a booklet listing the music broadcasts of a broadcasting station. The results are given in Table III.

[35] Frank Stanton, "Notes on the Validity of Mail Questionnaire Returns," *Jour. Applied Psychol.*, 23:95-104, 1939, p. 102.

[36] *Ibid.*

[37] E. A. Suchman and B. McCandless, "Who Answers Questionnaires?" *Jour. Applied Psychol.*, 24:758-69, 1940.

[38] *Ibid.*, p. 759.

TABLE III[39]

Proportion of Replies to Successive Waves of a Questionnaire on Radio Broadcasts

Wave	*Questionnaires sent*	*Percent returned*	*Cumulative percent of total*
First	820[a]	44.3	44.3
Second	457	46.4	70.2
Third	245	50.3	84.7
Fourth	122	66.5	95.1

(a) Eighty of the original 900 were eliminated from the sample on account of various types of unavailability. See *op. cit.*, p. 762, footnote.

The amount of follow-up necessary in any given inquiry will vary greatly with such factors as (1) the mandatory power, prestige, and personal relations of the investigator to the public investigated, (2) the simplicity of the questionnaire, (3) the educational level of the group solicited, and above all with (4) the interest of the public solicited in the subject under investigation. (5) Time and resources of the investigator and the degree of certainty needed in the particular inquiry will also determine how far to carry the follow-up.

K. WHO ANSWERS QUESTIONNAIRES?

When a questionnaire has been sent out, a certain number will be answered quite soon, others will trail in for a period of some months, and some will not be returned at all.[40] One question of great significance from the standpoint of the value of the conclusions is: Are the questionnaires returned representative of the results that would have been secured if all had been returned?

The purpose of many of the suggestions in this chapter has been to secure a high percentage of returns. The experience we have reviewed in the preceding section shows that

[39] *Ibid.*, p. 763.

[40] For a tabulation of how one commercial questionnaire to teachers was returned in terms of percentages at five-day intervals over a period of 122 days, see F. Stanton, "Validity of Questionnaire Returns," *Jour. Applied Psychol.*, 23:98, 1939. See also, E. A. Suchman and B. McCandless, "Who Answers Questionnaires?" *Jour. Applied Psychol.*, 24:758-69, 1940.

people vary in the amount of stimuli required to cause them to respond to questionnaires. The results actually achieved in different studies vary so widely and are conditioned upon so many variables as to make any generalization of very dubious value. We refer here to questionnaires which the recipient may ignore without penalty, and not official inquiries like the income tax blanks, or other questionnaires with coercive powers behind them. Most ordinary studies as conducted by private and relatively unskilled persons yield only from 5 to 20 percent of returns.[41] Certainly no reliable generalization can be based on such returns, because, as we shall see below, the sample is practically certain to be biased. That is, people who do not answer are likely to be different from those who do, also in other respects relevant to the inquiry. Even when a very high percentage of returns is secured, this bias is likely to be present. For example, Sletto[42] secured 69 percent returns from 1381 former university students on a 52-page questionnaire which had been very carefully prepared and pretested. Even in this large sample, it was found that 75 percent of those who failed to return the questionnaire were non-graduates, as contrasted with 60 percent non-graduates in the whole group. This suggests that even when the proportion of returns is high, biased results are quite possible. In short, a high percentage of returns is not in itself a guarantee of adequacy. As we have noted in Chapter V (p. 151), public opinion polls have brought to light some striking data regarding the selective workings of questionnaires in this field.[43]

From the experience of competent and large-scale users of questionnaires we may summarize the present state of knowledge on the subject of the selectivity of returns as follows:

1. Non-readers and non-writers are automatically excluded from participation. How seriously this may affect a study depends, of course, upon the public it is desired to reach. It must be remembered, however, that for purposes of an-

[41] F. Stanton, *op. cit.*, p. 102.

[42] R. F. Sletto, *op. cit.*; see also, R. C. Pace, "Factors Influencing Questionnaire Returns from Former University Students," *Jour. Applied Psychol.*, 23: 388-97, 1939.

[43] Elmo Roper, "Sampling Public Opinion," *Jour. Amer. Statis. Asso.*, 35: 325-34, 1940.

swering questionnaires the census definition of illiteracy is much too simple. Perhaps 25 percent of the general adult population of the United States is illiterate for questionnaire purposes.[44]

2. Interest in, or familiarity with, the topic or product under investigation is unquestionably a major factor in determining the return of a questionnaire (Stanton,[45] Watson,[46] Suchman and McCandless[47]). Non-users of a product answer in much smaller proportions, whether from a lack of interest or a reluctance to admit non-use (Salisbury[48]). This effect has been noticed especially in connection with personal and luxury products.[49]

3. The better educated are, generally speaking, more likely to return questionnaires than the less educated (Suchman and McCandless[50]).

4. Articulateness (measured by number of words written in response to "free answer" questions) is correlated to some extent with intellectual training but not with characteristics such as age, income level, size of community, marital status, or occupation of husband. Differences in articulateness are associated with differences in interests, which may themselves be the result of differences in amount of education. For example, a study of preferences in radio programs showed that the less articulate prefer needlework, dramatic programs, while the more articulate prefer reading and classical music.[51]

5. Follow-up procedures are important in decreasing the bias of the returns from the original request, because of the selective influences noted above. Follow-up procedures are therefore preferable to a mere increase in the number of original questionnaires in order to secure a return of a certain size, since the larger number would be likely to have

[44] See H. R. Huse, *The Illiteracy of the Literate*, (Appleton-Century, 1933), Chap. 1.
[45] *Op. cit.*
[46] R. Watson, "Investigations by Mail," *Market Research*, 7:11-17, 1937.
[47] *Op. cit.*
[48] Philip Salisbury, "Eighteen Elements of Danger in Making Mail Surveys," *Sales Management*, 34:28, 30, 84, 85, Feb. 15, 1938.
[49] *Ibid.*
[50] *Op. cit.*
[51] Isabelle Wagner, "Articulate and Inarticulate Replies to Questionnaires," *Jour. Applied Psychol.*, 23:104-15, 1939.

the same biases as the smaller number (Suchman and McCandless,[52] Rollins [53]).

6. The supposed greater economy of questionnaire studies as compared with interviews is seriously questioned by some authorities, if the slowness of securing returns (at least three weeks) and the cost *per return* is considered.[54]

L. CONCLUSION

Experience with mail questionnaires has shown that carelessly prepared schedules distributed to the general public yield from 5 to 20 percent returns. We have shown also that on account of the variation in the tendency to respond to questionnaires, generalizations from such results are practically worthless. In fact, they may be worse than worthless, because they frequently create the impression among the unwary that the subject has been studied "scientifically" and that the results have a validity which they clearly do not have. Such results are actually of little more consequence than those of the traveler and journalist who makes a trip abroad or from coast to coast talking with "lots of people," a taxi driver here, a grocer there, and usually finding that "public opinion" is practically unanimously in favor of his own viewpoint.

On the other hand, we have shown that when a questionnaire is carefully prepared according to the principles we have discussed, it is an instrument capable of yielding reliable results. The principal points to which we have directed attention in this connection are:

(1) A skillful appeal for cooperation.
(2) The enclosure of a stamped envelope for reply.
(3) Care in the wording and structure of the schedule.
(4) Its esthetic appeal.
(5) Adequate pretesting for

[52] *Op. cit.*

[53] M. G. Rollins, "A Practical Use of Repeated Questionnaire Waves," *Jour. Applied Psychol.*, 24:770-72, 1940.

[54] Frank Stanton, *op. cit.*, says, "When one considers the necessary wait that is a part of mail surveys — and the cost per return in addition to all other shortcomings — the advantages of the personal-interview technique far outweigh whatever advantages may be cited for the mail procedure" (p. 99). It should be noted also that the *Fortune* and the Gallup surveys of opinion both rely on interviewing rather than on mail questionnaires.

(*a*) its capacity to evoke response,
(*b*) reliability, and
(*c*) validity.

(6) Revision upon the basis of these tests.
(7) Sufficient follow-up effort to insure the representativeness of the replies received.

We have been concerned in the present chapter with mail questionnaires asking for information of any kind, from personal factual data, such as age and sex, to questions about beliefs, opinions, and attitudes. In the next chapter we shall consider in greater detail some instruments for securing the latter type of information. This will involve, in part, a more intensive consideration of a subject on which we have already touched, namely, the wording of inquiries or other forms of verbal stimuli to which we wish to secure responses. We shall be primarily concerned, however, with questions of measurement of attitudes through specially constructed tests and scales, by means of which the responses can be reduced to scores lending themselves to more refined analysis than is possible with ordinary questionnaire replies.

M. SUGGESTIONS FOR FURTHER STUDY

1. An excellent brief discussion of a careful experiment with questionnaires will be found in R. F. Sletto, "Pretesting of Questionnaires," *Amer. Sociol. Rev.*, 5:193-200, 1940. The student will find it worth his while to read this report in full.

2. Two complete numbers of *The Journal of Applied Psychology* (Vol. 23, February, 1939, and Vol. 24, December, 1940), are devoted entirely to the techniques and results of recent research on the responses of audiences to radio broadcasting. Several excellent papers on experience with questionnaires as well as other techniques will be found in the mentioned numbers of this periodical.

3. On the use of questionnaires in market research, see *The Technique of Marketing Research,* (McGraw-Hill, 1937), Chaps. 3, 4, 11, and 15, by P. F. Lazarsfeld.

4. On the wording of questions, an excellent discussion

will be found in G. Gallup and S. F. Rae, *The Pulse of Democracy,* (Simon and Schuster, 1940), Chap. 7.

5. For an excellent review of the kinds of bias that may creep into public opinion polls, see a statement by Professor Ross Stagner in the *Congressional Record* for May 9, 1941, pp. 3923-25. For a more detailed statement by the same author, together with a reply by G. Gallup, see "A Comparison of the Gallup and the *Fortune* Polls Regarding American Intervention Policy," *Sociometry,* 4 : Sept. 1941.

CHAPTER VIII

THE MEASUREMENT OF ATTITUDES AND OPINIONS

I see no more drawbacks in the path of setting up a branch of government whose job it is to determine by proven methods of selective sampling what the people want than there were in the path of setting up any of the existing branches of government. . . If such a branch were set up it would obviously have to be run by people with the same high ideals and sense of detached objectivity which are supposed to and in the minds of many do characterize the Supreme Court. I see no reason to suppose that a man automatically loses his integrity and his sense of objectivity when he is appointed to high office. Such a department could act as a sounding board for the public to inform its elected representatives more fully as to their desires than their ballot box vote could do. The public would then have an opportunity to inform its elected representative that he was elected because of the virtues of his planks 1 through 6, and despite the vices of plank 7 and thus admonish him not to regard his election as a mandate for putting this undesirable plank 7 into law. — *Elmo Roper.*[1]

A. PRACTICAL ASPECTS OF THE MEASUREMENT OF ATTITUDES AND OPINIONS

Whether one agrees with the above statement or not, it should serve to emphasize the practical implications of the interest in the measurement of attitudes and opinions. Attempts to gauge attitudes and opinions are as old as society itself. In small primary groups this measurement could be carried on with sufficient reliability through informal personal contact. With the development of rapid means of communication and the consequent interdependence of remote groups in a national or an international society, the demand for more formal and systematic methods of measuring opinions has arisen. Political and social leaders rely on various techniques under these circumstances. Some maintain a network of political agents reaching into every local community. In some places these function openly in their avowed capacity, or even as a type of social worker. In other places, they take the form of the secret police. Under relatively stable conditions these functionaries tend to be supplemented by a vast array of self-appointed journalists, newspapermen and propagandists who undertake to "reflect"

[1] "Sampling Public Opinion," *Jour. Amer. Statis. Asso.*, 35:333, 1940.

opinion. All of them are engaged in estimating, reporting, and trying to influence opinions and attitudes.

B. DEFINITIONS OF ATTITUDES AND OPINIONS

We are interested in this chapter not so much in the practical aspects of attitude and opinion measurement as in the scientific problem of how to observe reliably and to generalize from a subtle type of social behavior. The behavior which we define as attitudinal or as *an attitude* is a certain observable "set" of the organism or a reaction tendency preparatory to, and indicative of, more complete adjustment. In the present discussion, we shall use the term "attitude" to denote the general set of the organism as a whole toward an object or situation which calls for adjustment.[2] As such, it includes all the neural and other physiological sets and postures toward a situation, and their psychological correlates, commonly designated by such terms as "inclinations, feelings, prejudice, bias, notions, ideas, fears, threats and convictions. Thus a man's attitude about pacifism means here all that he feels and thinks about peace and war,"[3] as evidenced by some *objectively observable behavior* primarily of a preparatory, symbolic, or partially effective nature.[4] In short, studies of attitudes are concerned with the so-called subjective side of human behavior. (See Chapter I, Sec. G.)

For convenience in discussion it is well to distinguish two types of attitudes: (1) The various overt bodily postures or muscular tensions preparatory to some overt adjustment such as, for example, the readiness for defense, flight, grasping objects, etc. (2) The emotional or intellectual ("inner," "psychic") attitudes, which are observable (a) through our own kinesthetic and subvocal responses, and (b) through the symbolic behavior — mainly language — of others.[5]

[2] *Cf.* L. L. Bernard, *Social Psychology*, (Holt, 1928), p. 246.

[3] L. L. Thurstone, "Attitudes can be Measured," *Amer. Jour. Sociol.*, 33:529-54, 1928. For an excellent analysis of the hopeless confusion in current usage of the term "attitude," see Read Bain, "An Attitude on Attitudes," *Amer. Jour. Sociol.*, 33:940-47, 1928.

[4] Bernard, *op. cit.*, p. 246. G. W. Allport gives the following definition: "An attitude is a mental and neural state of readiness, organized through experience, exerting a directive or dynamic influence upon the individual's response to all objects and situations with which it is related." *Handbook of Social Psychology*, (Edited by C. Murchison, Clark Univ. Press, 1935), p. 810.

[5] *Ibid.*, p. 247.

Undoubtedly, one difficulty in attitude measurement is the variety of phenomena the term is used to denote. Symonds[6] records seven ways in which the term attitude is employed in current literature, viz., (1) great organic drives, (2) muscular set, (3) generalized conduct, (4) neural set or readiness to adjust, (5) emotional concomitant of action, (6) feeling concomitant of action, (7) accepting or rejecting verbal responses. Insofar as any or all of these phrases describe objectively observable phenomena they are legitimate subjects for study and measurement. However, our techniques of observing, classifying, and generalizing in the case of some of these different types of data are more highly developed than in the case of others. The type which falls perhaps peculiarly within the domain of the social sciences is that of verbal behavior, (Class 7 in Symonds' list). For the most part, this behavior takes the form of *opinions*. The present chapter, therefore, is chiefly concerned with methods of measuring verbal behavior which takes the form of opinions and their possible correlation with other forms of behavior.

LaPiere[7] has helped clarify the nature of attitudinal behavior by the following classification of all human behavior into four phases or levels:

(1) Overt-symbolic, which includes the acts of speaking, writing, and gesturing; (2) overt-nonsymbolic, which includes such directly significant acts as driving a car and closing a door; (3) covert-symbolic, or what is commonly designated as thought; (4) covert-nonsymbolic, or what is usually described as feeling-states and emotions.

Class (1) is clearly the phase with which most attitude studies hitherto have been concerned, and will be our chief concern in this chapter. Class (2) refers to a phase of behavior which most students would probably declare non-attitudinal. Classes (3) and (4) are perhaps generally regarded as attitudinal, but are by some considered not subject to measurement. We shall consider this question in the next section.

[6] P. M. Symonds, "What is an Attitude?" *Psych. Bull.*, 24:200-01, 1927. See also, G. W. Allport, *op. cit.*, pp. 806-10, for an excellent summary of present uses of the word "attitude."

[7] R. T. LaPiere, "The Sociological Significance of Measurable Attitudes," *Amer. Sociol. Rev.*, 3:175-82, 1938.

The study of attitudes, as defined above, is important in three ways. First, it is important from the standpoint of successful and peaceful living with other people. Obviously it is of great importance to know the emotional and intellectual behavior-pattern, ideologies, etc., of a group before formulating any plan for motivation, organization, or control of the group. It is important "to be able to recognize the significant types of attitudes in an adjustment situation where it is necessary to guide or divert responses. The orator, actor, salesman, advertiser especially, must be acquainted with the language of attitudes if he wishes to gauge successfully and control adequately his efforts." [8] Second, a knowledge of the processes by which attitudes are built up is important from the standpoint of social control. "The social and educational psychologists desire to know how the attitudes are built up and what environmental pressures—educational and social—to bring to bear upon individuals and groups of individuals to build up the desired response mechanism." [9] There is obviously great practical usefulness in a reliable method of determining and of measuring attitudes, because it would give us a measure of predictability of concrete behavior "without going to the elaborateness of studying concrete life conditions, for after all, the ultimate aim of personality-measurement is certainly the devising of short-cuts to the prediction of human behavior, and thus to afford a means of controlling it." [10] Third, the principal method of defining *true* social groups may be to discover what population groups have common attitudes, rather than assuming that they constitute *social* groups merely because they live in the same geographic area, or political subdivision, or receive an equal income, or because of other criteria which may be largely irrelevant as indicators of true group boundaries or "fellow feeling." In short, the attitudes of a group, especially in the form of symbolic behavior, constitute a large and important part of the phenomena necessary to describe, explain, predict and control the behavior of that group.

[8] Bernard, *op. cit.* See also, D. Katz, "Attitude Measurement as a Method in Social Psychology," *Soc. Forces,* 15:479-82, 1937.

[9] Bernard, *op. cit.*, p. 249.

[10] Kimball Young, "The Measurement of Personal and Social Traits," *Publ. Amer. Sociol. Soc.,* 21:97, 1927.

C. THE MEASURABILITY OF ATTITUDES

It is the apparent intangibility of such phenomena as attitudes that has given rise to the feeling that such data are not observable or measurable. If so, this interposes a serious obstacle in the way of a very exact science of group behavior. But, as we have seen, more critical examination reveals that the difference between such phenomena and "physical" phenomena is more apparent than real.

An attitude as defined above is obviously a complex and many-sided phenomenon. To argue that it can be wholly described by a single numerical or other index would indeed be a dangerous contention. But as Thurstone has pointed out, "for the problem of measurement this statement is analogous to the observation that an ordinary table is a complex affair which cannot be wholly described by any single numerical index. So is a man such a complexity which cannot be wholly represented by a single index. Nevertheless we do not hesitate to say that we measure the table. . . The context usually implies what it is about the table that we propose to measure (*e.g.,* its height, its cost, or beauty, or degree of appropriateness, or the length of time required to make it). . . We say without hesitation that we measure a man when we take some anthropometric measurements of him. The context may well imply without explicit declaration what aspect of the man we are measuring, his cephalic index, his height or weight or what not."[11] In the same way, if we specify or imply by the context *what aspect* of people's attitudes we propose to measure, it is just as permissible for us to say that we are measuring attitudes as to say that we are measuring tables or men.

This view also meets the objections that have been raised to the idea of measuring covert-symbolic behavior, commonly designated as "thought," and covert-nonsymbolic behavior, usualy called "feeling-states" and "emotions." Take, for example, the statement: "The psychogalvanometer may reveal changes in a criminal's feeling-states, but what any person is actually feeling and actually thinking cannot be

[11] L. L. Thurstone, *op. cit.,* pp. 530, 531, 536.

measured and is still anybody's guess."[12] This statement seems to overlook the fact that no measurements ever measure *all* aspects of any phenomenon. Measurements aim only to record reliably a single aspect, namely, *the aspect that is specified in the construction and legitimate use of the measuring instrument employed.* If so, we may point out that if the psychogalvanometer "reveals changes" *consistently and reliably* in a criminal's feeling-states, the observable concomitant fluctuations in the two is of the essence of measurement.[13] The *significance* of this consistent fluctuation *for stated purposes,* and the further correlation of the psychogalvanometer readings with *still other* phenomena in which we are interested, are questions to be answered *in each case* by actual research *on that problem.* Likewise, the statement that "what any person is actually feeling and actually thinking . . . is still anybody's guess" is undoubtedly largely true, especially if by measurement of thought and feeling is meant that the measurer must actually himself experience the thoughts and feelings of the person he is measuring. But this is surely a preposterous requirement which could not be met by any measurement in any field.

For our purposes it is enough to point out that "guessing" ("judging," "estimating," etc.) what other people think and feel is a daily pursuit of almost everybody, and that some are very much better guessers than others. We are interested in discovering and improving still further the technique of good "guessing." We do not expect that any one technique will ever enable us to guess *all* that voters thought about a certain candidate. We are content to be able to "guess" only, for example, that on the basis of our measurement of what they thought on a certain subject, as revealed by certain verbal behavior, a certain proportion (within one-half of one percent of the actual proportion, for example) would vote a certain way on a certain day. It is on the basis of

[12] R. T. LaPiere, *op. cit.,* p. 176. It should not be assumed from this quotation that this excellent article is incompatible with the viewpoint of the present chapter. The main thesis of LaPiere's paper is the unjustifiability of assuming, in the absence of positive evidence, any *necessary* correlation between the four levels of behavior given on p. 213. This cannot be emphasized too strongly, and is treated in our next section.

[13] "What is here significant is that all comparison is of the nature of measurement." John Dewey, *Logic. The Theory of Inquiry,* (Holt, 1938), p. 202.

such mundane performances that the question of what is or is not measurable will be settled in the end. *A priori* declarations on what is or what is not measurable, will contribute nothing to knowledge. But science has assumed in the past *as a working hypothesis* that some things not measured at a given time are not necessarily immeasurable if the proper instruments are developed. Scientists must continue to work on this hypothesis.

Now the particular form of attitudinal behavior which interests us most from the standpoint of sociology is that which serves as a stimulus-response mechanism between people, and which significantly affects their social interaction and their adjustment to their environment. By far the greatest and most important part of this attitudinal behavior is in the form of language. In its more complete form this behavior takes the language form of *opinions*. The principal aspect of attitudes which we are interested in measuring, therefore, is that aspect which takes the form of opinions expressed in language.

D. THE RELATION OF VERBAL ATTITUDES TO OTHER BEHAVIOR

In how far is significant measurement of mere opinions or other language expressions an index of the general "set" of the organism? What we are ultimately interested in, it may be argued, is complete overt adjustment in a concrete situation, and this behavior is frequently and notoriously at variance with people's verbal professions and opinions. While it is true that one of our interests in attitudes is the clue they may afford to the complete adjustment of the individual or the group, this is not their only significance. Whether a group of people actually does or does not practice permanent monogamy, the fact that it verbally declares its belief in this practice is still an aspect of its social behavior which is of vast significance. Its verbal expressions on the subject become perhaps the most important pressure of the social environment enforcing the monogamic sex mores in the community. The measurement of opinions, therefore, is of itself a matter of importance.

The fact that opinions or verbal behavior are frequently very unreliable guides to what a person would do in a con-

crete situation, does not destroy the value of the verbal behavior as an expression of attitude. It has been suggested that a person's actions (meaning his more complete, overt, public, and non-verbal adjustments) are a safer indication of his attitude than what he says. But overt actions are frequently designed to distort or conceal private, covert attitudes quite as fully as public verbal behavior sometimes conceals private attitudes. When the politician kisses Negro babies as an expression of his friendly attitude toward the Negro race his action is probably no more an accurate index of his "private" attitude than is his impassioned declaration of friendship from the platform. Both are just as "true" and real attitudes *in their respective situations* as the views he expresses or the action he takes toward Negroes after election. The situation in each case is different. The point is that our knowledge of attitudes can come only through a study of behavior, and *all* behavior is subject to modification in the process of execution from considerations of courtesy, expediency, or other social pressures. The attitude which a person exhibits, whether through verbal or other behavior, in a given situation, represents his attitude *in that situation.* How his responses in a hypothetical or symbolic situation correlate with his behavior in a similar concrete situation involving complete overt adjustment can be determined only by a study of both kinds of behavior in their respective situations.[14] We can make the index reliable only through a careful study and analysis of the conditions under which the

[14] LaPiere, (*op. cit.*), gives striking illustrations of the lack of correlation between verbal and other behavior in some concrete situations. The student is urged to read this article as a caution against unwarranted assumptions from opinion questionnaires and other verbal tests. See also his "Attitudes *vs.* Actions," *Soc. Forces,* 13:230-37, 1934. Much confusion has arisen from the use of such terms as "true" and "real" attitudes to designate the relatively permanent, integrated, private, and covert attitudes, *i.e.,* what a man "thinks in his heart," as contrasted with "false" or "misleading attitudes exemplified by rationalizations, conventional opinions, and public expressions of attitude conforming to the mores. All attitudinal behavior is, as such, equally "true" and "real." That is, the attitudes expressed in pencil and paper tests are as "true" and "real" as what one thinks privately or what one does overtly in a specific situation. (For examples of the use of such terms as "real" and "true" attitudes, see the cases cited in G. W. Allport, *op. cit.,* p. 809.) The question of how these various behaviors correlate with each other is an important subject for study, but has no bearing on the "truth" or "reality" of any of the behaviors observed.

behavior takes place — the whole situation — and from this infer whatever neuro-psychic or neuro-muscular set of a relatively permanent kind which we choose to hypothesize and call the attitude. All scientific truth is inferred in this way from the observation of objective behavior.

The technique of determining degree of reliability of attitudinal behavior as an index to complete overt adjustment behavior in a concrete situation would consist of correlating the attitudinal behavior toward a situation with the complete overt adjustment behavior. Broadly speaking, we may say that verbal behavior is likely to be a reliable index of other behavior when the behavior in question is not strongly governed by the current mores. Replies to the question, "Do you like baseball?" are likely to be "true" because in the current American mores there is no taboo or any special virtue attaching to either a positive or a negative attitude or behavior on the subject. On the other hand, a question such as, "Do you favor extra-marital sex relations?" might yield results very much less "reliable" in the sense here under discussion. Chapin[15] has suggested that a correlation between (*a*) the speech reactions, (*b*) "total overt" behavior and (*c*) past (historical) behavior of a person or group would reveal the degree to which one or two of these factors are an index of the others. If such correlations were extensively carried out they might become the basis of legitimate generalizations as to the degree of correlation to be expected between two or more aspects of behavior relative to given situations. But in the meantime we should (1) make clear exactly what aspects of attitudinal behavior we expect to study, and (2) limit our generalizations to the type of behavior studied.

The term *attitude,* then, will be employed in the present discussion merely as a blanket symbol for a hypothetical "set of the organism toward the object or situation to which adjustment is called for"[16] or a reaction tendency "preparatory to overt adjustment."[17] But as Bernard points out, "it can

[15] F. S. Chapin, *Cultural Change,* (Century, 1928), pp. 433 ff.
[16] L. L. Bernard, *op. cit.,* p. 246.
[17] *Ibid.,* p. 248.

be known and communicated to another only through its overt symbolic responses."[18] Hence, the only scientific approach to the study of attitudes is the selection of some or all of these responses — speech, bodily postures, or others — and a description of them under given conditions.

With these considerations in mind, we may proceed to a consideration of the types of measurement that have been devised to measure attitudes.

E. METHODS OF STUDYING ATTITUDES

As indicated above, there are two general types of approach to a person's social attitudes: (1) We may study as far as possible all of a person's past and present overt non-verbal and verbal behavior in given situations, and from this behavior infer his attitude; or (2) we may infer his attitude from a study of his verbal behavior in the form of specific opinions.

We are constantly imputing certain attitudes to people on the basis of their general behavior. For example, we make inferences regarding a person's religious and moral attitudes from his habits of church-going, his reactions to profanity, the books in his library, the pictures in his home, etc. We may be doing so systematically and for scientific purposes, or merely for purposes of our own or his immediate needs of adjustment. In either case, the more complete our information of a person's past behavior and the conditions under which it took place, the more accurately we can gauge his attitudes. The principal types of sources of such data may be classified as follows:

1. Life History Documents (Chapter XI)
   (*a*) Biographies
   (*b*) Case Histories (Social work records)

[18] *Ibid.* There is, of course, no objection to employing the term "attitudes" to indicate subjective states or processes susceptible, with our present technique, only to introspective study. As such, the term may be valuable as an hypothetical concept and may have cognitive value. See Ellsworth Faris, "Attitudes and Behavior," *Amer. Jour. Sociol.*, 34:271-81, 1926. Also Pitirim Sorokin, *Contemporary Sociological Theories*, (Harper, 1928), pp. 655-59. Also C. H. Cooley, "The Roots of Social Knowledge," *Amer. Jour. Sociol.*, 32:59-80, 1926. Much of the controversy on this subject is due to a misunderstanding on the use of the terms "subjective" and "introspective." For an important discussion of the possible objectivity of the symbolic process and so-called subjective states see J. F. Markey, *The Symbolic Process*, (Harcourt, 1928), pp. 144-48. See also, G. A. Lundberg, *Foundations of Sociology*, (Macmillan, 1939), pp. 10-34.

(*c*) Letters
(*d*) Diaries or memoirs
2. Oral interviews (Chapter XI)
3. Questionnaires and polls (Chapter VII).

It should be observed that the above classes are not mutually exclusive. Life history documents, for example, might be secured through oral interviews or through questionnaires, as well as through collection of the documents as prepared for other purposes remote from any thought of their use as scientific data. In this respect, however, the case history as prepared and used by social agencies perhaps resembles a kind of inquiry comparable to that of the interview and the questionnaire rather than to that of letters, diaries, etc. The fact is that all sociological inquiry necessarily deals with segments of social living, and the above categories are merely crude designations of very general and not mutually exclusive ways of securing and analyzing samples of such life-segments. We shall devote ourselves in this chapter chiefly to methods of studying attitudes through the methods of polls, tests, and scales, and leave the other techniques to other chapters (Chapters VII, IX, X, XI).

F. ATTITUDE QUESTIONNAIRES AND POLLS

The preceding chapter dealt with questionnaires designed to determine matters of fact as well as matters of opinion. The principles there set down will also apply to many of the ways of determining attitudes and opinions to be discussed below. The extensive use of the "Yes," "No," "No opinion" type of question with variations to suit the type of inquiry in the national public opinion polls has been accompanied by careful study and pretesting of this type of attitude test. The experience has yielded many data regarding the value of questions of this type and the importance of care in their construction.[19] Although these questions are administered for the most part through interviews, the principles governing their formulation apply equally to either written or oral tests.

[19] See G. Gallup and S. F. Rae, *The Pulse of Democracy*, (Simon and Schuster, 1940), Chap. 7.

Examples of poll questions:

1. "Should college teachers be free to express their views on all subjects, including politics and religion? (May, 1936).[20]

Yes	59%
No	41%
	100%

2. "Do you approve of a married woman earning money in industry or business if she has a husband capable of supporting her?" (November, 1936).[21]

No	82%
Yes	18%
	100%

3. "To what social class in this country do you feel you belong? Middle class, upper class or lower class?" (April, 1939).[22]

Upper	6%
Middle	88%
Lower	6%
	100%

4. "Which way do you think relief should be given—in the form of work relief (such as a W.P.A. job), or as direct cash relief?" (May, 1939).[23]

Work relief	89%
Cash relief	11%
	100%

No attempt is made in this type of attitude study to express individual scores by a number. The results are interpreted entirely in terms of the number or percent of people responding in a particular way to a single statement, phrase, or any other attitude indicator. This method of study is advantageous and popular because of the ease and positiveness with which the results can be manipulated statistically, and the comparatively brief task it usually imposes upon the informant. Its chief disadvantage lies in the difficulty of formulating questions on which a true attitude can be expressed in an absolute and unqualified affirmative or negative. Actual attitudes tend to differ by infinitely small shad-

[20] *Ibid.*, p. 310.
[21] *Ibid.*, p. 309.
[22] *Ibid.*, p. 309.
[23] *Ibid.*, p. 295.

ings and gradations. Thus in such questions as the first three above, some people would hesitate to commit themselves entirely without reservation. One way of avoiding the need for reservations is to phrase the question in sufficiently extreme form so that most people could without reservation indicate their attitude by "Yes" or "No." To the extent that this remedy is resorted to, however, the purpose of the investigation is largely defeated. For we are usually interested in attitudes not on questions on which we already know the unanimity to be great, but on the finer gradations of attitudes. That is, we are interested chiefly in attitudes on "borderline" questions—questions on which there is much reasonable difference of opinion.[24] On such subjects it is frequently very difficult to formulate a question that can be answered with an unqualified "Yes" or "No."

To meet the above objection, a third category, "doubtful," is frequently provided as a third possible answer. In order to afford more adequate opportunity to express gradations of attitude in this type of study various forms of arbitrary scales have been invented.

G. ARBITRARY SCALES

The devices for measuring attitudes to be considered in this section are called arbitrary scales because the distances between the steps or units depend on arbitrary assumption rather than on experimental determination. Some arbitrary rules of scoring are also usually adopted in these scales so that responses of different individuals are reduced to a certain abstract numerical score. These scores in relation to each other represent the relative degrees to which each person is supposed to hold a given attitude. It is only when they are so scored that they constitute a scale. The answers to many of the questions given below, for example, could be handled statistically as in the case of the polls considered above. That is, the percentage of a population answering a certain way could be computed. If so, these same tests would belong in the previous classification rather than in the class of scales.

[24] See Stuart A. Rice, "The Political Vote as a Frequency Distribution," *Jour. Amer. Statis. Asso.*, 19: (N.S.), 70-75, 1924.

Four types of arbitrary scales, each with a number of subtypes, can be distinguished and illustrated as follows:

1. *Point scales.* These are characterized by the fact that, in scoring, one point is usually given for each agreement or disagreement with a statement, or words, or situations as described or pictured. The following is an example of a common type:

The cross-out test. *Example:*

Directions: Read through the words listed below. Consider each one quickly. If it suggests more that is disagreeable than is agreeable, cross it out. You may cross out many or few words. Work as rapidly as you can, but be sure you cross out every word which is more annoying than pleasing, more agonizing than appealing, more distasteful than attractive.

Bolshevist	Dancing
Mystic	Prayer
Sunday Blue Laws	Birth Control
Ku Klux Klan	I. W. W.

(The complete test has 51 words.)

This type of test illustrates what is perhaps the simplest method of ascertaining attitudes through verbal tests.[25] The method is based on the principle that a tendency to cross out an unusually large number of words as annoying or distasteful is an indication of a general emotional set, or at least indicates an attitude on the symbolized subjects. To the extent that the simplicity of this test secures a spontaneous reaction, and to the extent that the conditions under which it is given are such as to reduce to a minimum the tendency of the person tested to be influenced by considerations of what others will think if he responds entirely spontaneously, this method of discovering attitudes undoubtedly has some degree of validity. The latter condition can perhaps be best approached by giving the test in a large group, and by not requiring the person taking the test to attach his name or other identification mark to the paper.

[25] The above illustration, from G. B. Watson, *The Measurement of Fair-Mindedness,* (Teachers College, Columbia University, 1925), also contains data on the reliability and the validity of the test. See Chap. 4 and pp. 89-91. For other and more recent examples of its use, see R. Stagner, "Measuring Relationships among Group Opinions," *Public Opinion Quarterly,* 2:622-27, 1938. Also R. Stagner, "The Cross Out Technique as a Method in Public Opinion Analysis," *Jour. Soc. Psychol.,* 11:79-90, 1940.

Two considerations affect the validity of the results achieved by this type of test. First, it is not always easy to select a list of words sufficiently suggestive of at least two specific attitudes. Many thoughtful individuals will find it difficult to do themselves justice when confronted with the list because they may wish for certain reasons to cross a word out, and for certain other reasons to leave it unchecked. When a large number of words of the same general implication or meaning are crossed out, however, it is probably a safe, though rough, indication of unfavorable attitude toward the things suggested by the words crossed out. Second, in many situations, attitudes cannot be expressed in a single word or phrase.

Stagner has used the test recently in studying configurations or associations between attitudes. From a test of forty items including such words as *labor union, socialist, capitalist, spiritualist, Holy Communion,* and other phrases identifying social, religious, and political institutions or ideologies, tetrachoric correlations were computed showing to what degree the dislike of a particular idea like *capitalist* is associated with dislike of such words as *Republican, Wall Street,* etc. Because of the limitations noted above, however, the cross-out test is not among the more frequently used tests.

Examples of another "point scale" type of attitude test are the more frequently used lists of statements in which the score is simply the sum of the endorsed statements,[26] and others which are scored by summing both the endorsed and rejected items in the test.[27] In the latter case, the statements are first classified into broad categories, such as "radical" and "conservative." The score is the sum of all the conservative statements endorsed plus all the radical statements with which the respondent disagrees. All of the "point" scales are useful for some purposes, although they make no attempt to register the gradations in intensity of attitude which some of the more refined scales attempt.

2. *Ranking scales.* These consist of a series of words (*e.g.*, nationalities, occupations, etc.), statements, or pictures of peo-

[26] C. Kirkpatrick, "A Methodological Analysis of Feminism in Relation to Marital Adjustment," *Amer. Sociol. Rev.*, 4:325-34, 1939.

[27] T. F. Lenz, "Generality and Specificity of Conservatism-Radicalism," *Jour. Educ. Psychol.*, 29:540-46, 1938. Also E. Nelson, "Radicalism-Conservatism in Student Attitudes," Psychol. Monograph., 50, No. 4, 1938.

ple or situations. The persons whose attitudes are being determined are asked to rank the material in order of preference as indicative of their attitude.[28] The results may be scored in various ways. Usually ranking tests are scored according to the degree to which the responses agree with a predetermined assignment of attitude value to each item. The responses may also be scored according to the degree to which they deviate in a significant manner from chance expectancy.

Example: Horowitz[29] tested racial attitudes by securing twelve photographs, eight of Negro children and four of white children who had been judged for racial typicality and general pleasantness. The photographs were presented to school children — kindergarten through the eighth grade — with the instruction: "Pick out the one you like best, next best, next best . . ." and so on until the pictures were all ranked. Scoring was by summing the ranks assigned the 4 white faces. If these faces were ranked 1, 2, 3, 4, the score was $1+2+3+4=10$; if they were ranked 1, 2, 5, 9, the score was 17. The lower the score, the greater the preference for the white faces. The possible range was from 10 to 42. If the ranking had been done by chance, the score might be expected to be 26. Deviation from the chance expectancy may be taken as indicative of the operation of a bias."[30]

As a variation of the above, the children were asked to select from the same twelve photographs companions for a variety of imagined situations. For twelve different situations the children were asked such questions as the following:

"1. Show me all those that you want to sit next to you on a street car
4. Show me all those that you want to come to your party
6. Show me all those that you want to go home with you for lunch

[28] With reference to the optimum number of ranks to ask for, Hazel Gaudet ("The Favorite Radio Program," *Jour. Applied Psychol.*, 23:115-26, 1939), expresses the opinion that three choices yield the best results. The number probably varies according to the subject-matter. See also, G. B. Hotchkiss and B. Franken, *The Measurement of Advertising Effects,* (Harper, 1927), pp. 38-40.

[29] E. L. Horowitz, *The Development of Attitude Toward the Negro* (Archives of Psychology, Columbia University, New York, 1936).

[30] *Ibid.*, p. 10.

7. Show me all those that you want to sit next to in the movies
11. Show me all those that you want to live next door to you

"Scoring the Show Me test was by finding the relative frequency with which the white faces were selected for all activities, expressed as a percent. The number of companions selected for all twelve items of the test was determined, then the number of those selections which were of the white boys (maximum, of course, 48); the frequency of the white choices was then computed as a percent of the total number of selections. Since of twelve faces, four were white, if the selections of companions were on a chance basis, the score might be expected to be $33\frac{1}{3}\%$. Deviation from the chance expectancy may be taken as indicative of the operation of a bias.

"On these two tests, the interpretation of the scores is this: on the Ranks test scores below 26 show preference for white as compared with Negroes (the smaller the score, the greater the preference); on the Show Me test scores above $33\frac{1}{3}\%$ show prejudice for white as compared with Negroes (the higher the score, the greater the prejudice against Negroes)." [31]

Words, phrases, and statements rather than pictures can be ranked, of course, and the results scored in the same manner. Indeed, the use of pictures is the exception rather than the most common practice.[32] The use of photographs or motion pictures of people and situations may be of great value in groups to which spoken or written language is of doubtful meaning. On the other hand, such tests are relatively difficult and expensive to administer and, while the stimulus is non-verbal, the response is necessarily verbal, as in written tests. In the case of children or groups of mixed and varied languages, these tests have obvious value either alone or as supplements to other methods.

[31] *Ibid.*, pp. 9, 10.

[32] For examples of ranking scales of the verbal type, see Helen Pallister, "Vocational Preferences of School-Learners in a Scottish Industrial Area," *Brit. Jour. Psychol.*, 29:144-66, 1938. Also A. L. Porterfield, "Education and Race Attitudes," *Sociol. and Soc. Res.*, 21:538-43, 1937.

3. *Rating (intensity) scales.*[33] In this type of instrument, steps on an attitude continuum are represented by responses (such as "strongly agree," "agree," "disagree," "strongly disagree") to separate statements or other attitude indicators not necessarily related to each other or to the same subject. The statements constituting these tests are frequently identical with those used in public opinion polls calling for responses of "Yes," "No," "Doubtful," or "No opinion." Indeed, if the answers to the statements on an intensity scale are scored simply in terms of the percentage of a population responding to each degree of intensity, the rating type of arbitrary scale should be classified rather as a type of opinion questionnaire or poll. When, however, arbitrary weights are assigned in scoring to each degree of intensity, such as 1, 2, 3, 4, 5 in the example below, each statement becomes an arbitrary scale.

Examples of intensity scales:

(*a*) Directions: Please indicate your opinion about each of the statements below by drawing a circle around the letter or letters in the margin which express your judgment. This is what the letters mean:

T-True (absolutely)
PT-Probably or partly true
D-In Doubt, Divided, Open Question
PF-Probably or partly False
F-False (absolutely)

T PT D PF F 11. The United States is justified in keeping the Philippines in order to insure our supply of raw rubber.

T PT D PF F 18. Japan sells more goods to the United States than to any other foreign country.

T PT D PF F 26. It is not justifiable for a loyal American citizen to protest against any course of action followed deliberately by the United States Minister to China.[34]

(*b*) In the interest of permanent peace, we should be willing

[33] For a good detailed discussion, see J. P. Guilford, *Psychometric Methods*, (McGraw-Hill, 1936), Chap. 9.

[34] This illustration is taken from the schedule of the Institute of Pacific Relations, 129 E. 52nd St., New York City.

to arbitrate absolutely all differences with other nations which we cannot readily settle by diplomacy.[35]

Strongly approve	Approve	Undecided	Disapprove	Strongly disapprove

(*c*) In order to aid you in expressing your opinion more clearly, you will find below a scale which will enable you to *check your degree of liking* from +10, meaning "exceptionally good," to −10, meaning "very poor," with zero representing an average amount of like or dislike.

"Referring to each program named above, would you please place a *check on the scale at the point you feel represents your degree of liking for the program?* Please keep the same order: Scale 'a' corresponds to program 'a'." [36]

Very poor		Average		Exceptionally good
/		/		/
−10	−5	0	+5	+10

In example (*a*) the first two statements are in less desirable form than the third. The first is confusing to many people in that it seems to contain two propositions: (1) that the United States is justified in keeping the Philippines; and (2) that the maintenance of our rubber supply is a valid reason for keeping the Philippines. When confronted with this type of statement people will frequently say: "I believe we are justified in keeping the Philippines, but not for this reason. How shall I answer?" If the point on which opinion is desired is the validity of the reason stated for our retention of the Philippines, as is clear from the statement as a whole, the possible ambiguity in interpretation mentioned above could be avoided by rephrasing the proposition so as to emphasize this; thus: "The maintenance of the raw rubber supply of the United States justifies us in keeping the Philippines."

The second statement under (*a*) is objectionable because it is a matter of statistical information rather than of opinion or attitude. To be sure, the question may be of value in show-

[35] From G. Murphy and R. Likert, *Public Opinion and the Individual*, (Harper, 1938), p. 19. The special type of scoring developed by these authors will be discussed in a later connection.

[36] Hazel Gaudet, "The Favorite Radio Program," *Jour. Applied Psychol.*, 23: 120, 1939. See also, H. E. Pemberton, "Optimum Rating Scale for Public Opinion," *Sociol. and Soc. Res.*, 17:470-72, 1932.

ing how exaggerated is a person's idea of the importance of our commercial relations with Japan. But since most people know that this is a question on which rather accurate objective information is available, they will tend to decline to express themselves unless they happen to know definitely. In short, answers to this type of question tend to reveal information rather than the less objective attitude which we are trying to measure in attitude questionnaires. As a question on information, it is not, of course, objectionable.

The last of the three questions under (*a*) above comes closer to revealing an attitude in the usual sense. It involves a question of personal attitude toward ethical or patriotic ideals, and as such falls more closely into the realm which attitude investigations are designed to survey.

This type of test aims to discover people's attitudes by their tendency to hold extreme or positive views on doubtful issues. Its value will, therefore, depend largely on the care with which the statements are selected. They should be statements which are decidedly doubtful, impossible to verify objectively, and on which reasonable people may disagree. If this is true of the questions used, then an extreme view in either direction may properly be regarded as indicative of definite attitude or bias. Only the expression of the extreme view in either direction, therefore, is usually scored as indicative of definite attitudinal bias.

This type of test has the advantage over the more arbitrary "True" or "False," "Yes" or "No," type in that it allows for a gradation of opinion or attitude. The results therefore permit a tabulation of the attitudes with respect to *degree of positiveness* as well as to their negative or positive character relative to a situation. This is a matter of considerable importance in drawing conclusions from the results, and is further emphasized in some of the examples that follow.

The number of possible responses in intensity scales may, of course, vary from two to six or more. On this point one study [37] revealed (*a*) that more people were willing to respond when a four-step response was permitted; (*b*) that the re-

[37] E. E. Ghiselli, "All or None versus Graded Response Questionnaires," *Jour. Applied Psych.*, 23:405 ff., 1939.

quirement of absolute choice between two mutually exclusive statements made the expressed opinions distinctly less favorable on the issue tested (belief in the sincerity of advertising); and (*c*) that the "yes-no" type of responses is unreliable as a measure of average opinion. Day reports that in current literature on the subject, seventy-seven percent of the scales are of the five-response type.[38] Again, the stimulus to the response may be a general theoretical statement, a pictorial behavior situation or a verbal behavior situation.[39] The behavior situation test may be illustrated as follows:

1. You are reading in a public library. A Negro comes in and sits down beside you.
   a. You rise and go to another table.
   b. You keep right on with your reading unconcerned.
   c. You complain to the librarian.
   d. You leave the library at once.[40]

2. Your state needs an additional tax so that its budget can be balanced. If you could vote, and a tax bill which provided for a widely distributed general retail sales tax was submitted to the people of the state for approval, what would you do?
   a. Not vote at all.
   b. Vote for the bill.
   c. Vote against the bill.
   d. Vote for the bill and try to persuade others to vote for it too.
   e. Vote against the bill and try to persuade others to vote against it too.[41]

There is some evidence that attitudes expressed toward concrete behavior situations have a higher reliability than the-

[38] Daniel Day, "Methods in Attitude Research," *Amer. Sociol. Rev.*, 5:395-410, 1940. Some recent examples of the five-response type of scale may be found in: M. M. Smith, "Comparative Social Attitudes," *Jour. Educ. Psychol.*, 28:681-93, 1938; F. C. Sumner and A. S. Campbell, "Attitudes Toward the Administration of Justice," *Jour. Psychol.*, 8:23-52, 1939; E. G. Williamson and J. G. Darley, "The Measurement of Social Attitudes of College Students: I. Standardization of Tests and the Results of a Survey," *Jour. Soc. Psychol.*, 8:219-29, 1937; C. N. Winslow, "A Study of the Extent of Agreement Between Friends' Opinions and their Ability to Estimate the Opinions of Each Other," *Ibid.*, 8:433-42, 1937.

[39] For examples of behavior situation scales see Leigh Peck and Elma Beckman, "An Experimental Study of Children's Attitudes Toward Relief," *Jour. Soc. Psych.*, 8:205-18, 1937; A. C. Rosander, "An Attitude Scale Based upon Behavior Situations," *Jour. Soc. Psych.*, 8:3-15, 1937; and R. N. Ford, "Scaling Experience by a Multiple-Response Technique: A Study of White-Negro Contacts," *Amer. Sociol. Rev.*, 6:9-23, 1941.

[40] A. C. Rosander, *op. cit.*

[41] C. R. Pace, "A Situation Test to Measure Social-Political-Economic Attitudes," *Jour. Soc. Psychol.*, 10:331-44, 1939.

oretical and opinion types of questions.[42] It is also held that a question involving a behavior situation is likely to be more valid as an index of actual behavior in such situations because it merely asks the direct and simple question, "What would you do?" This is perhaps true to the highest extent when the situation described is one which the responder has actually encountered, or which is not too fantastic or too foreign to his experience.

4. *Statement or unequal interval scales.* These scales consist of statements arranged by one or a few persons in what appears to them to be a logical order from favorable to unfavorable or some similar characterization. The similarity of these tests to the public opinion poll and the ranking method on the one hand, and the experimental or equal interval scales on the other hand, will be noted. The statement scale, however, is usually scored by the scale number (which has been assigned in advance) of the checked statement or the central tendency of the scale numbers of two or more endorsed statements. In this way the statement scale differs from the statements of the polls and the ranking scale, which have not been assigned ranks or scale values in advance. The difference between the statement scale and the experimental scales, to be discussed below, lies entirely, as we shall see, in the manner of selecting and arranging the statements. Several types of statement or unequal-interval scales are in common use.

Examples:

(*a*) Directions: In each of the following you are given a preliminary statement which can be completed or answered in any one of five ways. Check whichever one of the five most closely approximates your own opinion.

4. Income taxes should be levied:

a. Only upon those with immense fortunes.
b. Only upon those having ample means.
c. Only upon the upper and middle classes.
d. Upon all except the very poor.
e. Upon all.[43]

[42] D. D. Day and O. F. Quackenbush, "Relations Between War Attitudes and Opinions," *Sociol. and Soc. Res.*, 25:19-27, 1940.
[43] G. Murphy and R. Likert, *op. cit.*, p. 17.

(*b*) The Social Distance Test.[44]

According to my first feeling reactions I would willingly admit members of each racial group (as a class and not the best I have known, nor the worst members) to one or more of the groups under which I have placed a cross (X). (If you are wholly unfamiliar with any one of the racial groups, then no marks need be made. Note that crosses may be put in any number of the first five columns for each racial group. Note that a cross put in either row 6 or 7 for a given racial group means that no other crosses should be given that group.)

	Armenians	*Bulgarians*	*Canadians*	*Chinese, etc.*
1. To close kinship by marriage.			×	
2. To my club as personal chums.			×	
3. To my street as neighbors.			×	
4. To employment in my occupation.			×	
5. To citizenship in my country.			×	
6. As visitors only to my country.	×	×		
7. Would exclude from my country.				

This device is merely a variation of the rating system involved in the illustration immediately preceding. It is designed to measure the "social distance" between the subject and various nationalities or races. The same device can, of course, be used in relation to occupations, interests, and other social conditions. Its chief difficulty lies in the construction of the scale representing degrees of "distance" or other factors measured. As it is, it represents a valuable method of measuring relative attitudes.[44a]

[44] Note that if the method of equal appearing intervals, to be described below, has been applied to the statements in the social distance test, the test should be classified as an experimental or equal-interval scale. For an example, see below, p. 242.

[44a] The example given above is from E. S. Bogardus, *The New Social Research*, (Miller, Los Angeles, 1926), p. 212. For a more recent treatment of this subject by the same author, see "Scales in Social Research," *Sociol. and Soc. Res.*, 24:69-75, 1939.

H. EXPERIMENTAL SCALES

These scales differ from the arbitrary scales discussed above in that the statements or other attitude indicators are sorted or separated by a large number of judges in such a way as to furnish a rational, experimentally determined basis for the attitude continuum which constitutes the scale. Three forms are in current use.

1. The method of *paired comparisons.* In this method every item toward which an attitude is to be indicated is compared with every other item, one of the two being ranked above the other in each pair. The following example will illustrate the form of this test:

Example:

This is an experimental study of group attitudes. You are asked merely to underline the one of each pair with whom you would rather associate. For example, the first pair is

Englishman — South American *

* The complete test contains similar pairs of every nationality paired with every other, care being taken that no nationality is repeated twice in succession or in any other rhythmical pattern, and that every race appears an equal number of times on the right and the left of the pair.

If, in general, you prefer to associate with Englishmen rather than with South Americans, underline *Englishman.* If you prefer, in general, to associate with South Americans, underline *South American.* If you find it difficult to decide for any pair, simply underline one of them anyway. Be sure to underline one of each pair even if you have to guess.[45]

Calculation of scale values is based on the proportion of persons choosing one particular object or statement of attitude as against its alternative. These judgments with reference to a given object or statement tend to distribute themselves according to the normal probability curve, giving for each item its *discriminal dispersion.* The standard deviation of this dispersion is the unit of the scale constructed on the basis of the judgments of all the pairs. The full details of the method of constructing such a scale cannot be given here, as the method is laborious and requires considerable space for adequate elaboration. A lucid exposition of the subject, together with several variations and labor-saving

[45] H. Meltzer, "Group Differences in Nationality and Race Preferences of Children," *Sociometry,* 2:88, 1939.

devices, will be found in J. P. Guilford's excellent text.[46] The method of paired comparisons shows high reliability and validity as conventionally estimated and, with the exception of the method of equal appearing intervals, must be regarded at present as the most rational and reliable of all techniques of constructing attitude scales. As a matter of fact, the results of these methods are highly similar, except that in the method of paired comparisons items having a scale difference of an irrelevant nature may remain undetected.[47] It is true that in many cases equally reliable results can be secured through less laborious methods, and that for some purposes less accurate methods may be sufficiently good. The fact remains that we frequently do not know how generally this similarity of result secured by the simpler methods can be depended upon to occur. Scales constructed by the method of paired comparisons may serve, therefore, as valuable criteria of validity against which to check less accurate and less dependable methods.

2. The *order of merit* method.[48] This type of scale is identical with the ranking method as described above under arbitrary scales, except that the results are treated as in the case of the method of paired comparisons. A computational short-cut in determining scale values has been prepared by Barnhart.[49]

3. The method of equal-appearing intervals (*equal-i method).*[50] This is on logical, rational and theoretical grounds the most defensible method of measuring attitudes yet developed.[51] The method was first developed by Delboeuf for the measurement of sensory discrimination,[52] but its ap-

[46] *Psychometric Methods,* (McGraw-Hill, 1936), Chap. 7. See also H. Gurnee and E. Baker, "The Social Distance of Some Common Social Relationships," *Jour. Abnormal and Soc. Psychol.,* 33:265-69, 1938. This study applies the method of paired comparisons to the social distance test.

[47] L. W. Ferguson, "The Requirements of an Adequate Attitude Scale," *Psychol. Bull.,* 36:665-73, 1939.

[48] An excellent discussion of the details of procedure in this method will be found in J. P. Guilford, *op. cit.,* Chap. 8.

[49] E. N. Barnhart, "A Computational Short Cut in Determining Scale Values for Ranking Items," *Psychometrika,* 4:241-42, 1939.

[50] For elaboration of this method and detailed methods of computation, see J. P. Guilford, *Psychometric Methods,* Chap. 5.

[51] For a review of the evidence for this statement, see L. W. Ferguson, *op. cit.*

[52] See E. G. Boring, *History of Experimental Psychology* (Appleton-Century, 1929).

plication to attitude measurement is primarily the work of L. L. Thurstone, whose work in the field is of such dominance and distinction that scales of this type are quite generally, and justly, called "Thurstone scales." Thurstone's own description of the method of their construction can probably not be improved upon:

> Several groups of people are asked to write out their opinions on the issue in question, and the literature is searched for suitable brief statements that may serve the purposes of the scale. By editing such material a list of from 100 to 150 statements is prepared expressive of attitudes covering as far as possible all gradations from one end of the scale to the other. It is sometimes necessary to give special attention to the neutral statements. If a random collection of statements of opinion should fail to produce neutral statements, there is some danger that the scale will break in two parts. The whole range of attitudes must be fairly well covered, as far as one can tell by preliminary inspection, in order to insure that there will be overlapping in the rank orders of different readers throughout the scale.
>
> In making the initial list of statements several practical criteria are applied in the first editing work. Some of the important criteria are as follows: (1) the statements should be as brief as possible so as not to fatigue the subjects who are asked to read the whole list. (2) The statements should be such that they can be indorsed or rejected in accordance with their agreement or disagreement with the attitude of the reader. Some statements in a random sample will be so phrased that the reader can express no definite indorsement or rejection of them. (3) Every statement should be such that acceptance or rejection of the statement does indicate something regarding the reader's attitude about the issue in question. If, for example, the statement is made that war is an incentive to inventive genius, the acceptance or rejection of it really does not say anything regarding the reader's pacifistic or militaristic tendencies. He may regard the statement as an unquestioned fact and simply indorse it as a fact, in which case his answer has not revealed anything concerning his own attitude on the issue in question. However, only the conspicuous examples of this effect should be eliminated by inspection, because an objective criterion is available for detecting such statements so that their elimination from the scale will be automatic. Personal judgment should be minimized as far as possible in this type of work. (4) Double-barreled statement should be avoided except possibly as examples of neutrality when better neutral statements do not seem to be readily available. Double-barreled statements tend to have a high ambiguity. (5) One must insure that at least

a fair majority of the statements really belong on the attitude variable that is to be measured. If a small number of irrelevant statements should be either intentionally or unintentionally left in the series, they will be automatically eliminated by an objective criterion, but the criterion will not be successful unless the majority of the statements are clearly a part of the stipulated variable.

When the original list has been edited with these factors in mind, there will be perhaps 80 to 100 statements to be actually scaled. These statements are then mimeographed on small cards, one statement to each card. Two or three hundred subjects are asked to arrange the statements in eleven piles ranging from opinions most strongly affirmative to those most strongly negative. The detailed instructions will be published with the description of the separate scales. The task is essentially to sort out the small cards into eleven piles so that they *seem* to be fairly evenly spaced or graded. Only the two ends and the middle pile are labeled. The middle pile is indicated for neutral opinions. The reader must decide for each statement which of five subjective degrees of affirmation or five subjective degrees of negation is implied in the statement or whether it is a neutral opinion.

When such sorting has been completed by two or three hundred readers, a diagram like Figure 2 is prepared. We shall dis-

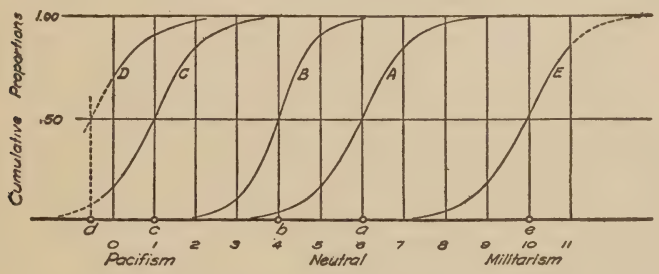


FIGURE 2. (From Thurstone.)

cuss it with the scale for pacifism-militarism as an example. On the base line of this diagram are represented the eleven apparently equal steps of the attitude variable. The neutral interval is the interval 5 to 6, the most pacifistic interval from 0 to 1, and the most militaristic interval from 10 to 11. This diagram is fictitious and is drawn to show the principle involved. Curve *A* is drawn to show the manner in which one of the statements might be classified by the three hundred readers. It is not classified by anyone below the value of 3, half of the readers classify it below the value 6, and all of them classify it below the value 9. The scale value of the statement is that scale value below which just one half of the readers place it. In other words, the scale

value assigned to the statement is so chosen that one half of the readers consider it more militaristic and one half of them consider it less militaristic than the scale value assigned. The numerical calculation of the scale value is similar to the calculation of the limen by the phi-gamma hypothesis in psychophysical measurement.[53]

The above description does not cover, of course, all the details in the construction of a Thurstone scale. For a full account the original monograph should be consulted.[54]

4. Mixed Methods. As has been indicated in the course of the above discussion of various methods, many tests are largely the same in both form and method of scoring. In some cases a change in the method of scoring a test changes its classification from an arbitrary to an experimental scale. For example, the simple ranking method becomes an experimental scale when the statistical methods of paired comparisons are applied to the rankings. Likewise the social distance test as usually constructed becomes an equal interval scale where the Thurstone method is applied to it. A few of the more significant combinations of methods should be mentioned in concluding this review of current methods.

a. Likert found that a great number of five-point arbitrary scales of the multiple choice, "statement," or "intensity" types yielded distributions resembling a normal curve. On this basis he devised the "sigma method" of scoring. The percentage of individuals that checked each of the possible gradations on a particular statement was converted into sigma (standard deviation) values. These values were taken from Thorndike's tables[55] showing the average values (in terms of standard deviations on a normal curve) represented by the stated percentages, the origin considered to be at the mean.[56] For example, if 13 percent check a certain statement "strongly approve," the Thorndike tables show that this corresponds to a sigma value of −1.63. Likewise, if 43

[53] For elaboration of this subject and methods of computation, see J. P. Guilford, *Psychometric Methods,* (McGraw-Hill, 1936), Chap. 6.

[54] For full details as to the construction of these tests, see L. L. Thurstone and E. J. Chave, *The Measurement of Attitude,* (Univ. of Chicago Press, 1929).

[55] E. L. Thorndike, *An Introduction to the Theory of Mental and Social Measurements,* 2nd Edition (Teachers College, Columbia University, 1913).

[56] For a full description of this method, see G. Murphy and R. Likert, *Public Opinion and the Individual* (Harper, 1938).

percent check "approve," 21 percent "undecided," 13 percent "disapprove," and 10 percent "strongly disapprove," the corresponding sigma values are, respectively, −.43, +.43, +.99, and +1.76. These sigma values now correspond to 1, 2, 3, 4 and 5 which are used ordinarily in scoring such series of statements. Observe that the sigma values, since they represent a percentage of people who have responded to a given statement in a certain way, constitute a kind of generalized scale in terms of which the responses to a large variety of questions can be compared.

This method has two conspicuous advantages, one of them chiefly practical, the other of theoretical importance. As compared with the Thurstone technique, the Likert method is much less laborious in that it eliminates the sorting of statements by a selected panel of judges and substitutes the voting or attitude expressions of the people actually taking the test. If the Thurstone scales are to be assumed to have fairly general validity, it must be assumed that the scale values, as arrived at, are independent of the attitude distribution of the judges who sort the statements, an assumption which has been questioned. Using the same statements as on a Thurstone scale, Likert's method has been found to yield a high correlation (in the cases tried) with the Thurstone scores, as was perhaps to have been expected. Likewise, the somewhat higher reliability shown by Likert's method (five possible responses) when applied to a Thurstone scale (two responses) is undoubtedly attributable to the demonstrated fact that as the number of possible responses increases (up to a certain limit) reliability also increases.[57] A scale which Likert constructed independently of Thurstone's method yielded correlations of only .65 and .71 with the latter.[58]

More extensive testing and comparisons of results of these methods are needed before final conclusions can be drawn. Basing the scale on equal-appearing differences to a random sample of the group to be measured rather than to an out-

[57] E. S. Conklin, "The Scale of Values Method for Studies in Genetic Psychology," *University of Oregon Publications*, 2:1-36, 1923.

[58] L. W. Ferguson, *op. cit.*, p. 671. See also, by the same author, "A Study of the Likert Technique of Attitude Scale Construction," *Jour. Soc. Psych.*, 13:51-57, 1941.

side panel of judges undoubtedly remains an important consideration in estimating the validity of the scale. All reported evidence, however, shows that scales constructed by the method of equal-appearing intervals, involving judges, are remarkably independent of the attitude distribution[59] of the judges. This is an important fact which has not yet been demonstrated for other types of scales.

b. R. N. Ford[60] combined some aspects of the Thurstone and Likert techniques in scaling statements of degrees of personal and community contacts with Negroes. Statements of such contacts, Ford found, could be scaled with reliability and validity according to the usual standards. While Ford's scales are ostensibly "experience" scales rather than attitude scales, they may be of great importance in checking our theories of the origins of attitudes and the relations between experience and attitude. Also, many scales of experience are likely to reflect attitude, as for example: "What has been your policy in the past when a Negro came into a public place, such as a theatre, railroad waiting room or restaurant?

1. Let the Negro know he wasn't wanted.
2. Let him alone so long as he knew his place.
3. Situation never arose.
4. Tried to pay no attention to the situation.
5. Treated him as I did whites in similar circumstances."[61]

Questions of personal contact and experience, when scaled, perhaps always tend to reflect attitude as well as mere experience. Questions of community contact, however, may be purely factual without revealing the responder's attitude. For example: "May a Negro run for election to public office in your home area?

5. Yes. Some hold or have held public office.
4. Yes. They may try.
3. Don't know of any such attempt.

[59] *Ibid.*, p. 667. Ferguson cites a number of studies which he asserts "have proved conclusively that scale values of items in a test constructed by the method of *equal-appearing intervals* are not affected by the attitudes of the judges. All reported correlations between the scale values of statements, based on groups of judges differing significantly in their attitudes toward the issues involved, are +.97 or over."

[60] R. N. Ford, "Scaling Experience by Multiple-Response Technique: A Study of White-Negro Contacts," *Amer. Sociol. Rev.*, 6:9-23, 1941.

[61] *Ibid.*, p. 14.

2. Legally, yes, but they never do.
1. Whites simply would not allow it." [62]

Ford had the group of five responses for each item sorted by seven judges according to the degree of friendship or hostility indicated. By comparisons of the ranking of the judges, ambiguous items were discovered and virtual unanimity of judgment was secured for the statements finally used. Ford's method is essentially a combination of the methods of Likert and Thurstone.

c. S. C. Dodd combined the Thurstone technique with the social distance test.[63]

In selecting statements for a social distance test the customary subjective judgment of the constructor of the test was replaced by the three criteria of (*a*) equal attitudinal distances,[64] (*b*) minimal ambiguity, and (*c*) maximal agreement between panels of judges. Thirty-nine statements, ranging from friendly to hostile, were rated on an eleven-point scale by sixty judges grouped in two panels. The resulting mean rated position of each statement was graphed, enabling the selecting of statements at equal intervals. The standard deviation (SD) of the sixty ratings of a statement measured its ambiguity. The difference [65] between the mean position assigned to a statement by the panel of students from all the Near East and by the panel of business men of Beirut is a measure of the degree to which the judges' rankings might depend on occupation, nationality, religion, age, and other factors which differed between the two panels of judges. A correlation of .99 was found between the thirty-nine positions assigned by the two panels. The five statements with the smallest differences between panels were then determined from the candidates remaining after applying criteria (*a*) and (*b*). The statements were assigned scores from 0 to 100 to serve as units of social distance expressed in percentages of the maximal distance, 50 being the neutrality point between friendliness and hostility.

The application of these tests to the conglomeration of religious, educational, national and economic groups in the Near East, and Dodd's analysis of the results is a fascinating

[62] *Ibid.*, p. 14.

[63] "A Social Distance Test in the Near East," *Amer. Jour. Sociol.*, 41:194-204, 1935. The method of paired comparisons was applied to the social distance test by H. Gurnee and Elizabeth Baker, "The Social Distances of Some Common Social Relationships," *Jour. Abnormal and Soc. Psychol.*, 33:255-69, 1938.

[64] See Thurstone and Chave, *The Measurement of Attitudes* (Univ. of Chicago Press, 1929).

[65] Expressed in units of the standard deviation of the difference.

DATA ON STATEMENTS ACCEPTED

Statement	*Mean Rated Position (11-point Scale)*	*Standard Deviation*	*Score Assigned*
A. If I wanted to marry, I would marry one of them.	.25	.44	0
B. I would be willing to have as a guest for a meal.	2.6	1.46	25
C. I prefer to have merely as an acquaintance to whom one talks on meeting in the street.	4.9	1.26	50
D. I do not enjoy the companionship of these people.	7.4	1.00	75
E. I wish someone would kill all these individuals.	9.6	.47	100
F. I know nothing about this group; I cannot express an attitude.	*	*	*

* Omitted from all calculations.

The test comprised four subtests adapted to the educational, economic, national, and religious groupings in the Near East. A contribution of this study was intended to be the measurement of social distances between such a diversity of groups by means of a single scale and type of unit.

example of this type of sociological analysis which can be made through attitude measurement.

I. ESTIMATING THE VALIDITY OF THE TESTS

As in the case of schedules and questionnaires in general, tests of the kind illustrated above need to be pretested and revised a large number of times before they are used for research purposes. Also, it is desirable to attempt to determine how closely the verdict of the tests agrees with other accepted criteria of existing attitudes. We shall return in the next chapter to a more thorough consideration of the value of social measuring instruments in general. The following general tests of validity suggest the type of checking to which attitude tests are subject.

1. The first and most subjective test of the validity of any device designed to measure attitudes lies in its logic and reasonableness with reference to tested existing knowledge. All

tests should, therefore, first be considered from the standpoint of how the principles on which they are constructed square with the established principles of science.

2. Where a number of tests are available, the correlation of the results of a particular test with the results of all of the other tests or of a group of them gives some presumptive basis for judging the particular test.

3. A third method of estimating the validity of a test is to secure a consensus of judgment from a group as to what individuals among their close personal friends hold pronounced attitudes on certain questions. The application of the test to these individuals then gives a measure of the degree to which the test agrees in its findings with the judgments of close associates. Individuals known for their objectivity of mind may also be confronted with their own scores on the tests and the degree of their acceptance of the results may be some indication of the validity of the test.

4. Certain groups may, in the judgment of competent persons, be expected to have certain biases. The ability of the tests to reveal these attitudes is some indication of the validity of the tests.

5. The degree to which intelligence, education, and familiarity with the language of a subject, rather than his attitudes, tends to be measured by the test may be determined by correlation of attitude test scores with scores of intelligence, reading tests, and other factors that are suspected of being influential.

6. In addition to these general tests of the validity of the tests given above, Thurstone suggests several objective measures of the validity of the attitude scale after it has been constructed.[66]

(*a*) *The scale must transcend the group measured.* — A measuring instrument must not be seriously affected in its measuring function by the object of measurement. . . If a yardstick measured differently because of the fact that it was a rug, a picture, or a piece of paper that was being measured, then to that extent the trustworthiness of that yardstick as a measuring device would be impaired. . . The assumption is . . . that two statements on a prohibition scale will be as easy or as difficult to discriminate for people who are "wet" as for those who are "dry" . . .

[66] L. L. Thurstone, *op. cit.*, pp. 547-52.

The experimental test for this assumption consists merely in constructing two scales for the same issue with the same set of statements. One of these scales will be constructed on the returns from several hundred readers of militaristic sympathies and the other scale will be constructed with the same statement on the returns from several hundred pacifists. If the scale values of the statement are practically the same in the two scales, then the validity of the method will be pretty well established.[67] It will still be necessary to use opinion scales with some discretion. Queer results might be obtained with the prohibition scale, for example, if it were presented in a country in which prohibition is not an issue.

(*b*) *An objective criterion of ambiguity.*—Inspection of the curves in Figure 2 reveals that some of the statements of the fictitious diagram are more ambiguous than others. The degree of ambiguity in a statement is immediately apparent, and in fact it can be definitely measured. The ambiguity of a statement is the standard deviation of the best fitting phi-gamma curve through the observed proportions. The steeper the curve, the smaller is the range of the scale over which it was classified by the readers and the clearer and more precise is the statement. The more gentle the slope of the curve, the more ambiguous is the statement. Thus of the two statements *A* and *B* in the fictitious diagram the statement *A* is the more ambiguous.

In case it should be found that the phi-gamma function does not well describe the curves of proportions in Figure 2, the degree of ambiguity may be measured without postulating that the proportions follow the phi-gamma function when plotted on the attitude scale. A simple method of measuring ambiguity would then be to determine the scale distance between the scale value at which the curve of proportions has an ordinate of .25 and the scale value at which the same curve has an ordinate of .75. The scale value of the statement itself can also be defined, without assuming the phi-gamma function, as that scale value at which the curve of proportions reaches .50. If no actual proportion is found at that value, the scale value of the statement may be interpolated between the experimental proportions immediately above and below the .50 level.

(*c*) *An objective criterion of irrelevance.*—In addition to these tests of validity, Thurstone gives a "criterion of irrelevance" of the statements used in the scale, which may be summarized as follows:

[67] "The neutrality point would not necessarily be represented by the same statement for both militarists and pacifists, but the scale separations between all pairs of statements should be practically the same for the two conditions of standardization." L. L. Thurstone, *op. cit.*, p. 548.

Assume that the whole list of a hundred statements on the subject of militarism-pacifism has been submitted to several hundred readers for their indorsement or rejection. Let these readers be asked to mark with a plus sign every statement which they endorse and with a minus sign every statement they reject. All of the statements have, it must be remembered, already been rated on the basis of the degree of pacifism they express and have been assigned a value in a frequency scale on the basis of the previous ratings. If a statement is relevant for our purpose, *i.e.*, if it is accepted or rejected primarily because of the attitude on pacifism which it portrays, then the rating of the other statements should arrange themselves with reference to it in a more or less normal frequency distribution. If on the other hand, the statement under consideration has implications which lead to acceptance or rejection for other reasons than the degree of pacifism which it conveys, *i.e.*, if it is irrelevant for our purpose, then the ratings of the other statements with reference to the basic statement will not fall in the form of a normal frequency distribution.

For example, assume that the hundred statements have been arranged as a frequency distribution in a scale from 0 to 8 on the basis of the degree of pacifism they represent in the opinion of a large number of readers. Assume that we desire to test the relevance of a statement which falls at point 4.0 in the scale. If the readers indorse or reject the statement chiefly on the basis of the degree of pacifism it represents, then these readers who indorse statements in the vicinity of 4.0 on the scale will not often indorse statements that are very far away from that point on the scale. Very few of them should indorse a statement scaled at point 8.0, for example. "If a large proportion who indorsed the statement scaled at 4.0 should also indorse a statement scaled at 8.0, then we should infer that their voting on these two statements had been influenced by factors other than the degree of pacifism that is implied in the statements."

We shall consider further in the next chapter certain basic questions of the validity of sociometric scales in general. In that connection we shall consider also the query: "*What* do sociological scales measure?"

J. SUMMARY AND CONCLUSION

We have described and illustrated above the principal methods currently used in attempting to measure a kind of human behavior. That behavior is called attitudinal behavior and was characterized in the opening sections of the present chapter. The strength and weakness of different methods has been noted in the course of the discussion. In conclusion, some considerations regarding attitude measurement in general should be mentioned.[68]

1. In the first place, the above examples of opinion questionnaires and tests in common use do not, of course, pretend to represent an exhaustive or complete list of such devices. Each general type has numerous variations to suit particular subjects or purposes. In addition to these relatively formal methods, we pointed out in our classification of methods (pp. 220-21) that illuminating studies of attitude have also been made through participant observer techniques, interviews, studies of voting behavior,[69] and the analysis of life history documents.[70] These will be considered in a later chapter.

The examples exhibited in this chapter are submitted as illustrative of the more formal and direct devices at present employed to determine certain stimulus-response patterns which exist in social groups. The behavior or intercommunication constituting these action patterns is the subject-matter of social psychology. As we have seen, this intercommunication is carried on chiefly by means of symbols — principally by language symbols. Hence we study these action-patterns — the attitudes — through the observation and interpretation of language responses, not only as verbal behavior, but as indications of more complete adjustments.

[68] Attention is again directed to two excellent summaries of the present status of attitude measurement by L. W. Ferguson and Daniel Day, to which frequent reference has already been made.

[69] See C. E. Robinson, *Straw Votes,* (Columbia University Press, 1932); G. Gallup and S. F. Rae, *The Pulse of Democracy,* (Simon and Schuster, 1940); L. Bean, *Ballot Behavior. A Study of Presidential Elections,* (American Council on Public Affairs, New York, 1940); H. F. Gosnell and Norman Pearson, "The Study of Voting Behavior by Correlational Techniques," *Amer. Sociol. Rev.,* 4:809-15, 1939.

[70] For an excellent appraisal of one such study, namely, W. I. Thomas and F. Znaniecki, *The Polish Peasant,* (Boston, 1918), see H. Blumer, *et al., Critiques of Research in the Social Sciences: I,* (Social Science Research Council, 1939).

2. It may be that a person is largely unable to analyze all aspects of his attitudes and opinions. But the analysis he does make is significant, as psychiatrists and others well know. At least his expressed attitude shows what he wants people to think he thinks. "I conclude . . ." says Cooley, "that the imaginations which people have of one another are the solid facts of society, and that to observe and interpret these must be the chief aim of sociology." [71]

We have emphasized, however, that the degree of correspondence between verbal behavior and other behavior is not of itself a conclusive criterion of the validity of the results secured through an attitude scale. To be sure, such correspondence is of interest and great practical significance. A great deal of the interest in attitude studies undoubtedly arises from a desire to predict also non-verbal behavior. But this should not cause us to overlook the fact that attitudinal behavior has its own significance. To be able to *predict attitudinal behavior* may be quite as significant as any other kind of prediction. The relationship of attitudinal to other behavior is, in fact, a separate problem.

3. It must always be remembered that responses to questionnaires and tests are speech responses to a total situation, and not only a response to the words of the questionnaire or to the hypothetical situation they represent. All the conditions under which the questionnaire is given are part of the situation. If the questionnaire calls for attitudes on subjects governed by strong mores, the subject's expressed attitude will tend to be the conventional attitude, especially if his name is required on the schedule.[72] That is, the subject will express what he *thinks* his attitude *ought to be* rather than what his behavior in the actual situation might be. As a verbal response his expressed attitude may still be of value and interest. But too many studies of attitudes assume that the verbal response must be a reliable index of other behavior. Discrepancies between the two can be partly eliminated by the form of the schedule, anonymity of the replies,

[71] C. H. Cooley, *Human Nature and the Social Order,* (Scribners, 1902), p. 87.

[72] See S. M. Corey, "Signed Versus Unsigned Attitude Questionnaires," *Jour. Educ. Psychol.,* 28:144-48, 1937. This author suggests on the basis of an empirical study that the significance of anonymity has been exaggerated.

and careful standardization of the conditions under which it is presented. If we knew enough about those conditions which prevent people from revealing their "true" attitudes in a given situation, we could measure the influence of disturbing factors and allow for them in our interpretation as does the physical scientist in analyzing and interpreting a situation outside the control of his laboratory. A correlation of "experience" or "situational" test results, not to mention everyday behavior, with the results of opinion and belief scales, is precisely what is needed in this respect. We have already referred to LaPiere's studies of this subject.[73] Where possible this adds greatly to the accurate interpretation of the results.

4. Full allowance must be made in the interpretation of attitude statements for the differences in language equipment, education, and general cultural background on the part of the subjects. Ambiguity of words and misunderstandings and misinterpretations on the part of the subjects may invalidate the replies as expressions of attitudes. These replies are, furthermore, subject to all the rationalizations and fantasies that play so large a part in thinking and in speech reactions. Antagonisms and inhibitions aroused by a question may prevent a frank expression of attitude.

5. The objection has been raised that the methods of measuring attitudes reviewed above are too specific in content to measure phenomena like attitudes, allegedly more general in character. Kirkpatrick[74] and Remmers[75] have accordingly constructed tests designed to allow for the more adequate expression of general attitudes. There seem to be no grounds for denying that attitudes of all degrees of generality exist, and that they are subject to measurement by essentially the same

[73] See R. T. LaPiere, *op. cit.*

[74] C. Kirkpatrick, "The Construction of a Belief-Pattern Scale for Measuring Attitudes Toward Feminism," *Jour. Soc. Psychol.*, 7:421-37, 1936; "The Content of a Scale for Measuring Attitudes Toward Feminism," *Sociol. and Social Res.*, 20:512-26, 1936; "Assumptions and Methods in Attitude Measurement," *Amer. Sociol. Rev.*, 1:25-88, 1936; H. Cantril, "General and Specific Attitudes," *Psychol. Monograph*, 42, No. 192, 1932.

[75] H. H. Remmers, "Studies in Higher Education," *Bull. of Purdue University*, 35, No. 4, 1934; H. H. Remmers and E. B. Silance, "Generalized Attitude Scales," *Jour. Soc. Psychol.*, 5:298-312, 1934. Some limitations of this type of scale are found by R. Stagner and N. Drought, "Measuring Children's Attitudes Toward Their Parents," *Jour. Educ. Psychol.*, 26:169-76, 1935.

types of techniques. Practically, therefore, this question reduces itself largely to what type of attitude interests us, and the types of items appropriate to the end sought.

From the theoretical point of view, it is well to remember in this connection that *no measurement of any phenomenon necessarily describes it completely.*[76] No scientific instrument pretends to measure all aspects or dimensions of the "objects" designated by folk language. It selects out only certain revelant aspects of an object or of behavior, and measures it. Now, one aspect of attitudes—the aspect we are interested in for our present purpose—is the "more or less" aspect of opinions relating to some subject. This is an aspect of people's opinions which we do measure empirically in our everyday contacts, when we say, for example, that one person is more radical or more conservative than another on some question, or that one person is more in favor of a specific proposal than another. All that the technology discussed in this chapter proposes to do is to make this procedure more objective and accurate.

6. Finally, the complete cultural relativity of all attitude measurements should be recognized. As Thurstone has pointed out, significant results could hardly be expected from a scale designed to measure attitudes toward prohibition in a country where prohibition is not an issue. Similarly, a scale to measure attitudes toward labor unions would obviously not yield significant results if applied to a non-industrial society. The possible influence of difference in cultural background must also be considered in much less extreme instances as, for example, when college people devise scales to be used on non-college populations. It is true that Thurstone scales have been found to be remarkably independent of the attitudes of the judges sorting the statements. But this is a subject which must be considered whenever attitude scales are used. It is not, however, a handicap unique to attitude scales. The readings of most physical measuring instruments must also be adjusted if these instruments are used under conditions varying widely from those for which

[76] This subject is also related to the questions of "what" attitude tests measure, the "nature" of "real" attitudes, etc. See Chap. IX, pp. 302-06 below. Also, G. A. Lundberg, *Foundations of Sociology*, pp. 61-65.

the instrument was constructed. An ordinary thermometer will give reliable readings only between certain temperatures and the readings of other instruments will vary, for example, with altitude. Techniques of detecting and allowing for such variations are, of course, an important aspect of research in attitude measurement.

7. Finally, there remain some technical and theoretical problems regarding the "nature" of the units used in attitude scales, and the consequent permissibility of certain mathematical operations with the scores secured.[77] These issues do not greatly affect the validity or demonstrated practical usefulness of the type of measurements discussed in this chapter as *instruments of discrimination,* but they are questions of theoretical and scientific significance that can be answered only by further research. The last decade has seen an enormous increase and improvement in attitude measurement, with a pronounced trend in favor of the equal-interval scales.[78]

[77] For a brief statement of these alleged limitations of equal-interval scales, see H. M. Johnson, "Some Neglected Principles in Aptitude-Testing," *Amer. Jour. Psychol.,* 47:159-65, 1935; "Pseudo-Mathematics in the Mental and Social Sciences," *Ibid.,* 48:342-51, 1936; C. Kirkpatrick, "Assumptions and Methods in Attitude Measurements," *Amer. Sociol. Rev.,* 1:75-88, 1936; R. K. Merton, "Fact and Factitiousness in Ethnic Opinionnaires," *Amer. Sociol. Rev.,* 5:13-28, 1940. For comments on Merton's paper, see remarks by G. A. Lundberg in the same journal, pp. 38-39, and by Jessie Bernard, same journal, pp. 415-17, together with a reply by Merton, pp. 647-48. A fuller statement of the present author's position will be found in G. A. Lundberg, *Foundations of Sociology,* (Macmillan, 1939), Chap. II. A somewhat different view will be found in M. R. Cohen and E. Nagel, *An Introduction to Logic and Scientific Method,* (Harcourt, 1934), Chap. 15.

[78] D. Day, (*op. cit.,* p. 400) gives the following summary:

TABLE IV

Comparison by number and percentage of attitude studies published in the periods 1925–1931 and 1937–1939 classified by methods employed

Methods	*Publications*			
	Number		*Percent*	
	1925–31	*1937–39*	*1925–31*	*1937–39*
Case	18	15	18.4	9.3
Absolute Ranking	31	78	31.6	48.1
Ranking Scales	12	6	12.2	3.7
Graphic Rating Scales	4	5	4.1	3.1
Statement Scales	19	6	19.4	3.7
Paired Comparisons	6	5	6.1	3.1
Equal-i Scales	8	47	8.2	29.0
Total	98	162	100.0	100.0

There is no question that great progress has been made toward a more refined technique of measuring a type of phenomenon with which the social sciences must be concerned to a large degree. There is great demand for reliable ways of gauging attitudes—for more accurate means of determining and predicting the social weather. Leaders and proponents of social programs of every kind always have been and always must be concerned with the attitudes of the group they want to influence. The technique thus far has been chiefly of the intuitive, rule-of-thumb type. An improvement is all that is claimed for the methods considered in this chapter. The future of this type of research must depend on still further improvements in these methods.

K. SUGGESTIONS FOR FURTHER STUDY

1. By far the best *brief* digest of the recent literature on the measurement of attitudes is Daniel Day, "Methods in Attitude Research," *Amer. Sociol. Rev.* 5:395-410, 1940. This excellent article contains an annotated bibliography of 200 items covering the literature which appeared in readily available books and periodicals during the years 1937–39 inclusive. An even more pointed and critical discussion of the subject is L. W. Ferguson, "The Requirements of an Adequate Attitude Scale," *Psychol. Bull.,* 36:665-73, 1939. The status of different types of scales from the point of view of the specified requirements, with a supporting bibliography of 45 items, is incisively discussed. For a more general discussion, see D. E. Scates, "The Essential Conditions of Measurement," *Psychometrika,* 2:27-34, 1937; and by the same author, "How Science Measures," *Jour. Exper. Ed.,* 5:296-312, 1937.

2. A somewhat longer discussion of methods, as well as a digest of studies and results, is Chap. 13 in *Experimental Social Psychology,* by G. and L. Murphy and T. Newcomb (Revised Edition, Harper, 1937). See especially the summary tables, pp. 948-51, 956-59, 982-85, 998-99.

3. The best general source on all types of scale construction, together with detailed methods of computing scale

values and scores, is J. P. Guilford, *Psychometric Methods* (McGraw-Hill, 1936). For detailed accounts of the techniques of construction of particular scales, see L. L. Thurstone and E. J. Chave, *The Measurement of Attitude* (University of Chicago Press, 1929); G. Murphy and R. Likert, *Public Opinion and the Individual,* (Harper, 1938), Chap. 3 and Appendix B; C. Kirkpatrick, "The Construction of a Belief-Pattern Scale for Measuring Attitudes Toward Feminism," *Jour. Soc. Psychol.,* 6:421-37, 1936; E. A. Rundquist and R. F. Sletto, *Personality in the Depression* (University of Minnesota Press, 1936); A. C. Rosander, "An Attitude Scale Based upon Behavior Situations," *Jour. Soc. Psychol.,* 8:3-15, 1937. See also, T. C. McCormick and R. C. Schmid, "A System of Attitude Experiments," *Soc. Forces,* 19:351-56, 1941. For a set of scales designed to measure to what extent motion picture plays violate the mores, see C. C. Peters, *Motion Pictures and Standards of Morality,* (Macmillan, 1933). See also, R. F. Sletto, *Construction of Personality Scales by the Method of Internal Consistency,* (The Sociological Press, Hanover, N. H., 1937).

4. References to methods of measuring a great variety of aspects of personality will be found in G. W. Allport, *Personality, A Psychological Interpretation,* (Holt 1937). A good general treatise is E. L. Thorndike, *The Psychology of Wants, Interests, and Attitudes* (D. Appleton-Century, 1935).

5. *A Bibliography of Mental Tests,* by Gertrude Hildreth (The Psychological Corporation, New York, 1939), contains 129 references to attitude tests.

6. For an important contribution toward a resolution of some controversial points regarding the types of scales that have been discussed in this chapter, see F. S. Chapin, "The Syntactical Analysis of Sociometric Techniques: Cases in Point." *Sociometry,* 4:177-83, 1941.

CHAPTER IX

THE MEASUREMENT OF INSTITUTIONAL BEHAVIOR

For practical reasons even more than for purely intellectual ones, we need rigorously scientific studies of human society and of our individual relations to it. In particular we need such studies of the societal interests that are labeled "public policy," "education," "missions," and "social work." I am aware that this proposition is resented by men and women who suffer from an anti-"academic" complex and worry lest the "human touch," "the ways of the neighborhood" and the naïve thinking of "plain people" shall have spontaneity squeezed out of them by theory. This is an unfortunate misapprehension of what science is and of what it does for us. It ought not to be necessary at this late day, but it is necessary, to tell the general public that science is nothing more nor less than getting at facts, and trying to understand them, and what science does for us is nothing more nor less than helping us to face the facts. — *F. H. Giddings.*[1]

. THE DEMAND FOR MEASUREMENTS OF INSTITUTIONAL BEHAVIOR

The appraisal of institutional behavior with reference to norms, either for comparison with other communities or with respect to ideals or ends toward which communities strive, has always been a matter of interest. In a small and homogeneous society of close personal contacts, the subjective, empirical, evaluative judgments of everyday life were accepted as adequate. But with the development of a large and heterogeneous society, there has come an increasing demand for evaluations and measures which shall be impersonal and shall have the general validity and objectivity of measures in the physical sciences, in business, and in industry. In education and psychology complicated intelligence and achievement tests have been developed which enable us to compare pupils as individuals and as groups with each other, with pupils in other schools, and with pupils in other cities, states, and countries. As long as our interest in the problem was confined to the one-room school with a dozen children, the teacher's subjective rating of the pupils served all the purposes for which ratings were desired. But today school executives demand ratings in terms that will enable comparisons with standards and norms far beyond the ex-

[1] *The Scientific Study of Human Society,* (University of North Carolina Press, 1924).

perience of the individual teacher. On the basis of such comparisons they draw conclusions regarding the efficiency, needs, and problems of different schools. Colleges and universities demand an evaluation of prospective students in *standard* terms. These ratings are taken into serious consideration in determining the disposal of individuals, whether as delinquents or as candidates for honors. Standard measures of "intelligence" and school achievement are becoming items of information in social diagnosis in a way comparable to such items as height and weight in a physical examination.

Obviously, the same needs which have given rise to these measures of the individual are also present when we consider groups and institutions. If we admit the importance of home environment in cases of juvenile delinquency, then an intelligent diagnosis of a case demands an index of such environment for the same general reason that we desire an intelligence quotient. Both aid us in accounting for the delinquency, and indicate the degree to which each factor may be responsible. Likewise a business man who is accustomed to insist on a rigid weighing and measuring in objective terms of money expended and returns received is increasingly inclined to demand an equally tangible accounting from the charitable, educational, or governmental institutions he is supporting.

We are concerned in the present chapter with methods of grading institutions, organizations, and communities. The types of such institutions and organizations are legion, and each type demands a different kind of yardstick, but the general principles for the construction of such yardsticks are the same for all. We shall consider, therefore, (1) the general requisites and desirable attributes of a valid sociometric scale ; (2) the concrete procedure in the development of such a scale, including the difficulties encountered in its construction. After considering the theory and principles of constructing sociometric scales, we shall conclude with a consideration of several concrete examples of them.

B. GENERAL REQUISITES OF A SOCIOMETRIC SCALE

The five principal general requisites of a desirable sociometric scale are as follows :

1. *It must be generally applicable, reliable, and valid.* A measuring instrument is considered reliable when it gives the same or highly similar results when repeatedly applied to the same or highly similar phenomena. A scale is called valid when it measures in some demonstrable way what it claims to measure. We have touched on these requirements in the preceding chapter and shall return to them again in a later section. By the general applicability of a scale we mean that it must remain reliable and valid when applied to any case within the range or class of phenomena for which the instrument was designed. How broad that class is to be depends entirely upon our purpose in constructing the scale. We may construct a scale measuring socioeconomic status of coal miners in Pennsylvania or of the entire population of the United States or even a more heterogeneous population. It is true that as a science matures it tends to develop some measures of extremely general validity, such as measures of length and weight. Even mature sciences, however, are constantly constructing special instruments to measure highly selected phenomena. These instruments sometimes have limited applicability.

How general an instrument should be in its application is a question to be decided with reference to the problem we wish to solve. The more generally applicable a scale can be made without impairing its reliability and validity, the better. Practically, there are usually limits beyond which it is not desirable to extend the applicability of any particular scale. If we strive for too great generality we may seriously impair the scale's validity as an intensive measure of a more restricted type or class of objects. For example, if we attempt to make a measuring scale for homes so general in its application that it will measure with validity the homes of Chinese coolies or of Australian aborigines as well as modern American homes, the requirements of such a scale might call for the inclusion of many items which would be necessary for the effective rating of one class of homes, but which would be totally irrelevant for another class. This might in turn seriously impair the value of the scale as an intensive measure of any of the classes. Likewise, the American Public Health Association, in its Appraisal Form for Public

Health Work which will be discussed later, found that a scale for rating the health departments of cities of 50,000 population, or over, would not be a fair measure of the health departments of smaller cities in many respects. For example, while the employment of a full-time health officer may be a reasonable point on which to score a city of over 50,000, this item would not necessarily be significant as a basis for grading in a smaller city, because (*a*) none of the smaller cities may have such provision, and (*b*) the presence or absence of a full-time official, in a small city, may not be an important index of the efficiency of health work in the city. While a scale should be generally applicable to the range of objects for which it is intended, we should avoid classifying together objects which have too little in common. It is better to limit the range of classification and to establish norms for each particular class.

2. *It must allow for and make possible the establishment of reasonable norms relative to which the phenomena measured can be compared.* This is largely a corollary of what has been said above regarding the classification of the objects or aspects to be measured. If we attempt to apply a scale to too wide a range of objects, the norms for some members of the class will be Utopian and largely worthless for practical purposes. One way to arrive at a norm is to take the statistical mode of a representative sample of the group to be measured. When it is not possible to establish a norm in this way, standards may be tentatively set by the combined judgments of experts or others regarded as best qualified to pass on the particular point under consideration. Methods of selecting such experts or authorities so that the results will command respect and obviate personal biases as far as possible, will be considered in a later section.

3. A third requisite of a good scale is that the *factors which go into its construction must be generally available.* It is frequently true in the present stage of development of the social sciences and with the present methods of record keeping in public institutions that many factors of great significance in determining the efficiency of an institution are not at present definable in terms sufficiently objective to permit comparison with corresponding factors elsewhere. In

such cases we cannot include such factors in a measuring scale, regardless of how important the factors may be. This is a difficulty which militates against the refinement and validity of all sociometric scales at the present time. We have not yet developed techniques for measuring many significant types of behavior. The development of such techniques is, as we have seen in previous chapters, the central problem of the social sciences. Until science makes available more objective definitions and measures of the significant factors in institutions we must utilize the factors and measures generally available. It avails us nothing in the construction of a practical measuring scale to require information in our scale which is not available in the great majority of cases which we propose to measure, regardless of how desirable that information may be. We must confine our analysis to factors available, although the resulting inadequacy will to some degree invalidate our scale as a reliable measuring device.

4. Within the limitations set forth above, a fourth requisite of a sociometric scale is that *it should take into consideration all significant aspects and factors for which data can be secured.* Here again we are confronted with the problem of defining significance, and of deciding which factors are of sufficient significance for our purpose to justify their inclusion in our scale. This problem involves considerations of units and other matters that will be discussed later. In the absence of means of statistical or other objective demonstration of the role played by each factor in the total functioning of the institution to be studied, we must again rely for *tentative* evaluation on the next best criterion, namely, on the combined judgment of a group of persons best qualified to judge. In the end these evaluations are tested by whatever criteria we finally accept as indicative of the validity of the scale as a whole, which we shall discuss later in this chapter.

We must have not only a selection of significant factors to include in our scale, but an evaluation of the degree of their significance. This leads us to the final requirement of a satisfactory scale.

5. *Each factor included in the scale must be weighted ac-*

cording to its relative significance. If we enumerate a dozen factors entering into the functioning of an institution, some will be of greater significance than others. A reliable scale must allow for this fact. The problem is largely the same as that involved in the original selection of the factors to be included. Both will be determined in the last analysis by the purpose of the scale, *i.e.,* what is to be measured or solved. In the absence of objective demonstration of the relative significance of different factors, we must again rely tentatively on the combined opinion of persons best qualified to judge. This is frankly a concession to necessity in the absence of scientifically developed norms. With the development of scientific study of social institutions there is no reason why even ethical judgments should not be based on relatively objective and verifiable data. In the meantime we are compelled to rely tentatively on our nearest approximation to scientific evaluation, namely, the combined judgment of persons in closest touch with the situation and whose habits of objective thought are generally recognized.

The application of these principles in specific cases will be illustrated below. With these general principles before us, we may now proceed to a consideration of the concrete procedure in the construction of a scale, including a consideration of the difficulties involved and the ways of meeting them.

C. PROBLEMS IN THE CONSTRUCTION OF A SOCIOMETRIC SCALE

The first step in the construction of a sociometric scale is the selection of the behavior, institution, organization, or community, aspects of which are to be measured. The unit to be measured may vary from some primary groups such as the family or neighborhood to the more secondary and derived groups, such as an industrial or business organization, the educational system as a whole or the public libraries alone, a department of the city government, the whole government, or the city as a whole. In addition to the selection of the unit to be measured, we must also decide what aspect of the unit we desire to measure. We might develop a scale designed to rate cities from the standpoint of their relative advantages for general industrial development; or for the

development of a particular industry; or as health centers. Again, we may design our scale to measure cities from a much broader viewpoint such as their general desirability for people of specified tastes, as a place in which to live. Other units are subject to the same variety of rating systems.

It is well to bear in mind that all measurement, whether of the solar system or of a grain of sand, is specific in that the measuring instrument *must specify what it takes into consideration. No* measurement describes *all aspects* of any natural phenomenon. To measure a tree usually means to measure its *height.* The frequency with which the term measurement in this case is used to designate the measurement of that particular aspect sometimes causes us to assume that in the other sciences "objects" "as a whole" are measured, and that in such cases the problem of what aspect to measure does not arise. But the situation is clearly not different in this respect from what confronts us in the measurement of social phenomena. To be sure, it it possible to combine measurements of many aspects into composite indices. But such indices, while "general" in their inclusiveness, are as *specific* in denotation as are their constituent factors.

Sociometric scales aim, of course, to discriminate accurately between different gradations of highly intricate complexes of social behavior. The scores assigned to the different cases measured may also serve as indices of a number of other social conditions with which the aspects measured happen to be reliably associated. The discovery of such correlations must, indeed, be a major task of the social sciences. The physician is able to tell a great deal, and to predict about a patient, from readings of a few simple instruments only because of well established correlations of this kind. Scales of the type here under discussion facilitate the establishment of such correlations and herein lies their larger scientific significance.

Two principal problems in the construction of sociometric scales, then, face us at the outset: (1) How shall we select the aspects or factors of a unit which we deem significant and which are therefore to be considered in our scale? (2) How shall we determine the relative weight to attach to each factor included?

D. THE FACTORS AND WEIGHTS IN A SOCIOMETRIC SCALE

The question as to what factors are significant in a situation is, of course, a recurrence of the problem which arises at the beginning of every investigation and which we considered in a previous chapter in our discussion of the working hypothesis. In all situations which have not been scientifically analyzed into the factors which can be demonstrated to operate in certain sequences, we assume such factors and sequences from intuition, analogy, logical deduction, or empirical observation. But in every situation, there are some people who are better qualified than others to select the significant factors in a specific situation. Such superior qualifications spring from closer contact with the facts in the case and training in logic and objectivity of mind. These are obviously the persons to make the original selection of the significant factors to include in a sociometric scale. We shall deal in a later section with methods of verifying these judgments.

The selection of a limited number of the best qualified persons for a particular purpose, however, is itself a problem. In the case of the sociometric scale we need two types of such persons to correspond to the two main problems mentioned above: (1) *General judges* to select the significant categories, or factors in a situation and to assign weights to them according to their relative significance. (2) *Special judges* to determine the criteria of efficiency of the functioning of each special factor selected by the general judges. The first class calls for persons of broad experience rather than of intensive specialization, persons who can see social institutions and processes as a whole and in proper perspective, and who can weigh the many interrelated elements with a high degree of objectivity. A group of representatives of the different social sciences would perhaps be the best qualified for this task. The second class of judges calls for specialists in each field selected as a factor in the unit to be measured — the physician, the banker, the engineer.

For example, if it is desired to construct a scale for measuring the relative desirability of cities as good places in which to live for people of conventional tastes, the first task is to

select the general departments or factors in a city which are considered of major significance from this point of view.[2] Who shall make this selection and determine the relative weight to be attached to each factor? Probably the best qualified person for such a task would be representative persons of the population at large. A group of economists, sociologists, political scientists, and city managers might be used, provided these professionals may be assumed to be familiar with the *public* taste, as perhaps contrasted with their own tastes, or their notions of what the public *ought* to want. Suppose that a committee consisting of such persons selects Health, Education, Public Administration, Recreation, and Industry, as the significant factors to be included. Another type of judge is obviously desirable to select and evaluate the important factors determining the efficiency of each of these departments. That is, a scale for the rating of health departments or educational systems would call for specialists in these respective fields—health administrators and educators.

The above division of judges into two classes on the basis of whether they are to determine the general factors to be considered and the weights to be applied to each, or whether they are to determine the methods of rating each general factor so selected, is, of course, itself a relative matter. Thus, in the example employed above, the persons charged with the duty of formulating a rating system for health departments would be regarded as special judges—public health experts. In formulating a scale for grading health departments, however, these judges would in turn assign special aspects or factors of public health work such as tuberculosis control, food, milk and water inspection, to the specialists in these subdivisions of the public health department. That is, specialists in tuberculosis control would formulate standards and methods of rating the effectiveness of the anti-tuberculosis organization. Experts in sanitation would do the same with reference to purity of foods, milk, and water, and the adequacy of the sewage system. But the relative importance to be assigned to tuberculosis control as compared to general

[2] E. L. Thorndike, *Your City*, (Harcourt, Brace & Co., 1939), has selected 37 such indices of "goodness of life for good people." See especially Chap. 2.

sanitation would have to be determined by the judges of health departments as wholes. Thus, with reference to the city as a unit, the administrators of health departments would be special judges. But with reference to particular aspects of public health work they would be general judges. For the rating of units larger than the city, likewise, the general judges for the selection of significant factors in the city as a whole might become special judges, relatively speaking. The distinction between general and special judges is, therefore, subject to extensive subdivision and is always relative.

The selection of the persons best qualified to construct a sociometric scale must, then, always be governed by the specific function they are to perform. Expertness may be defined in terms of all degrees of generality and specialization. The general superintendent of a large industrial plant may be regarded quite as properly a specialist and an expert with reference to his particular task, as is the most highly specialized scientist in the laboratory of the plant. The efficient plant will consist of a hierarchy of experts, each of which is best qualified to determine the profitable division of labor within his own sphere and, consequently, to pass on the qualifications of persons to perform increasingly specialized tasks. The problem of selecting the best-qualified persons for the construction of a sociometric scale of the type here under consideration resolves itself, therefore, largely into a question of selecting the general judges of the unit as a whole. After they have been chosen, the selection of judges of subordinate units must be accomplished by the generally accepted methods of determining the qualifications of persons in different fields.

Various methods of determining the relative eminence of men in the larger fields of human endeavor have been devised. One of the best known of these attempts is that by J. McKeen Cattell in his *American Men of Science.* In brief, Cattell's method was as follows: "In each of the twelve principal sciences the names were arranged in the order of merit by ten leading students of science. The average positions and the probable errors were then calculated so that in each science the order of merit was determined together with its validity. The names were then combined in one list by

interpolation, the number of workers in each science being taken approximately proportional to the total number of workers in that science." [3]

Cattell's system of selection, however, is adapted only to the better-established sciences. It is not generally applicable to the selection of all the types of persons who may be best qualified to pass judgment on certain aspects of a sociometric scale. A method more generally applicable is suggested by Gould.[4] Briefly, his procedure is as follows: (1) All the names in *Who's Who in America* were classified according to occupations, in so far as this was possible. A comparatively small number of names had to be omitted because of the ambiguity of the occupation stated. (2) The number of electors for each occupation was determined according to the number in the whole group. (3) Each of the electors was asked to rate in order of merit the twenty individuals in their occupational group whom he considered most eminent. While only ten judges were desired, twice this number were rated in order to provide substitutes for such of the first ten as were unable or unwilling to serve. (4) The rankings so secured were combined by adding the rankings given by all the electors to each eminent person in the respective fields. The ten persons in each field ranking the highest by this combined judgment were selected as judges in their particular specialty. In one of the occupational categories were classified the social scientists, community experts, general psychologists, physiologists, botanists, geographers, and geologists, as a group from which the best qualified persons might be drawn to serve as general judges of the sociometric scale. They were selected in the same way as the experts in more highly specialized fields.

Each profession and occupation usually has its own methods of determining the relative competency of its members, and these methods will ordinarily be employed in selecting persons to pass on the validity of the factors and weights employed in a sociometric scale. There are, furthermore, in many departments of social organization, national associa-

[3] J. McKeen Cattell, *American Men of Science,* (Science Press, 1906), Preface to the First Edition.

[4] K. M. Gould, *A Sociometric Scale for American Cities.* (M.A. Thesis, Columbia University, 1921.)

tions which devote themselves largely to the task of setting standards, such as the Child Welfare League of America, The National Child Labor Committee, the Children's Bureau of the federal government, the National Recreation Association, the American Public Health Association, and many others. The standards advocated by these organizations have usually been developed and tested by leading authorities. Whenever possible, their standards should be utilized in the construction of sociometric scales in preference to specially constructed norms. As such organizations develop they can test and revise further the standards necessary to develop reliable sociometric scales.

While the problem of selecting the best qualified persons for the construction of sociometric scales should be left as far as possible in the hands of organizations specializing in the field concerned, some general problems should be considered at this point, touching the construction of all scales, regardless of the subject matter which they are designed to measure. Among the most important of these problems is the type of units and indices desirable. Everything which was said in an earlier chapter regarding appropriateness, clarity, measurability and comparability of units applies to the selection of measuring units for sociometric scales. But some special applications of these principles should be mentioned.

Since all the significant aspects and factors of the unit measured cannot as a rule be included in a scale, factors should be selected which touch most directly and vitally the problems studied. Elmer has suggested that in the construction of a scale for most organizations three types of data are desirable:[5] (1) Population (demographic) data, or the relationship between the number of individuals who cooperate with or belong to the organization and the total eligible population of the community. (2) Institutional data, or measures of the degree to which the members participate in the work of the organization. (3) Functional data, or measures of the degree to which the purpose or purposes of the organization are being achieved. From the standpoint of the organization alone, the last of these measures would perhaps be sufficient. But from the standpoint of the community,

[5] M. C. Elmer, *Social Research,* (Prentice-Hall, 1939), pp. 298 ff.

the first is obviously important. The second may frequently be an index of the third. The relative weight to be assigned to each class of data will rest on experience and the combined judgment of the persons best qualified to judge the group concerned, as discussed above.

While these three categories are suggestive of the general types of data which are desirable, they are not literally applicable to all social units. For example, a public health department could not be scored in terms of the first type of data, namely, the proportion of the inhabitants who belonged to or participated in the organization. In such cases, however, this type of data is really supplied by information on per capita costs, per capita equipment, and personnel per unit of population. Likewise institutional data may be interpreted to include the training and experience of the directing force, and the presence of health machinery—hospitals, clinics, etc.[6] Functional data here, as in all cases, would, of course consist of the various indices of accomplishment—the extent of the activities to prevent sickness, the number of visits to clinics, the number of nursing visits, vaccinations, etc.

E. SUMMARY OF THE PRINCIPLES OF CONSTRUCTION OF SOCIOMETRIC SCALES

From these illustrations we may summarize the guiding principles in the construction of a sociometric scale as follows:

1. We should aim to select factors and units which measure the degree to which an institution achieves its purpose or purposes. In the last analysis, this must be the crucial test. Our ratings should reflect the extent to which members of an organization practice the principles which the organization exists to propagate. For example, to use Elmer's illustration,[7] the degree to which the members of a farmers' organization to promote crop rotation actually practice rotation of crops is the final test of the functioning of that organization. In the same way, the proportion of people who own their own homes, and the amount and distribution of mort-

[6] *Cf.* L. A. Boettiger, "The Wisconsin Better Cities Contest," *Soc. Forces*, 5:436-46, 1927.

[7] *Op. cit.*, p. 175.

gages, are more vital social indices for some purposes than the volume of building or per capita wealth. The tax rate is an important item, but it is relative to the goods and services furnished by the government. The qualifications, number, and pay of health officials, and the per capita expenditures for health are important indices, but the vital measure of the health department is the community's freedom from disease, its low death rate, if and when *traceable to the functioning of the department.* A large per capita park acreage may be a suggestive index. But not until we consider the *distribution* of that acreage, *i.e.,* the per capita acreage for smaller units than the whole city and relative to density of population and other factors, do we have a really vital social index. Likewise it may be that the average expenditure for police protection may be much lower in cities of 2500 than the average cost of such protection in cities of 100,000. But it does not necessarily follow that persons and property are safer in large cities because they spend, say, five times as much per capita for police protection. An index of per capita expenditure for police must be considered in relation to other indices—the amount of crime, arrests, etc.

2. Our insistence on functional units and indices does not mean that other indices are not important or that they are to be neglected. In many cases we can make a valid interpretation of the functional data only by considering the more static and frequently more objective institutional data. The mere fact that crop rotation is general in a certain area does not of itself prove that the farmers' organization is responsible or that this organization is functioning vigorously. Only when we find that a large percentage of farmers belong, that their meetings are frequent and well attended, and that the programs have devoted much time to the subject of crop rotation can we assume with some assurance that the organization is responsible for the result observed. The mere fact of a low morbidity and mortality rate does not establish the efficiency of a health department. But when we find a highly qualified and well-paid health department personnel, well organized and equipped, correlated with such low rates, *other conditions being constant,* the presumption of a causal connection between the two becomes stronger. Final assur-

ance as to causal connection must wait upon intensive comparative analysis of the relevant facts in different cities. Thus the demographic and structural data of a non-functional type are very important in three respects: (*a*) they are frequently the most available and objective; (*b*) when experience has shown that a high degree of correlation exists between the efficiency of an institution and the personnel qualifications, per capita costs, and other population ratios of a non-functional type, these data become in themselves reliable indices of the vital functional facts in which we are primarily interested; (*c*) demographic and structural facts are necessary for the proper interpretation of functional facts.

3. Whether the data called for in the scale are of the demographic, institutional, or functional type, there are certain principles governing the choice of units which apply equally to each type of data. For purposes of comparability, which is the essence of a scale, the units should be stated in terms of *rates, ratios, percentages,* or *standard index numbers.* Such ratios as the number of children in attendance per teacher, the library circulation per capita, the percentage of families owning their homes or engaged in paying for them, and specific death, birth, and marriage rates, are examples of the most useful types of units to be employed in a sociometric scale. Some of these units (*e.g.,* death and birth rates) should usually be averaged over a period of years, and the average employed for comparative purposes, rather than selected for the more or less immediate present. The period of time to use must be determined by the purpose of the particular scale or the particular investigation, and by the tendency of the data in a particular field to fluctuate.

4. We have already pointed out that the relative weights which are to be assigned to different factors in an activity must be left to the combined opinion of persons best qualified to judge. But these judges will be guided in their assignment of weights by the consideration of what aspects of the situation it is desired to stress. For example, if it is desired to emphasize the preventive rather than the curative aspects of a health department, public sanitation will be weighted more heavily than, say, the number of clinical visits or the hospital facilities. Likewise in the schedule em-

ployed in the Wisconsin Better Cities Contest, education was allowed 4000 points as against 1000 points for all other divisions because of the assumed preventive value of this aspect of community activity, and the desire to emphasize the value of preventive activities. Again, this schedule apparently aimed at emphasizing recognized standards of institutional organization rather than the functional aspects. So, out of a total of 190 points for the section on Family Social Work, 75 points were assigned to the question whether a case worker was employed "who has had a course in a regular school of social work, or in a university department of sociology, or who has had experience on the staff of a standard case work agency." The actual amount and efficiency of the family social work carried on by persons without these qualifications was scored relatively lightly or was ignored altogether. The assignment of weights, therefore, will be largely governed by our focus, our purposes.

In weighting, the question arises how to treat apparently incompatible indices. A high rate of taxation is burdensome, and from one point of view should operate unfavorably in the score of a community. Yet high taxation may be the means whereby exceptionally good schools, playgrounds, and health activities are maintained. Accordingly the numerical score will be largely determined by the relative weight we assign to these two aspects. In a case of this kind, both indices should be included as direct measures without reference to each other.[8] Each will then offset the other as far as the facts justify. If the tax rate is above the average, the score of the community will be lowered. But if it is getting full value for the expenditure of its public funds, its unfavorable score on taxation will be offset by its favorable score in other respects, to the extent that these are above the average.

5. In general, the same tests of validity of a scale which were discussed in the preceding chapter are applicable to the sociometric scale. (1) The simplest and crudest test is a comparison of the scale score with the empirical judgment of persons familiar with the units scored. This does not, of course, prove anything conclusively regarding either the scale

[8] M. C. Elmer, *op. cit.*, pp. 85, 86.

or the empirical judgments. A wide disparity in the two types of rating, however, may indicate the need for further inquiry into the validity of the scale. (2) The correlation between the scores of different scales applied to the same unit provides one kind of test of the validity of any or all of the scales thus correlated. (3) The general applicability of the scales can be determined only by application to a large number of units of considerable variety. For example, if we desire to test the validity of a sociometric scale for cities, we might apply it to a considerable number of cities of different sizes. We might then group the cities according to size and compare the average scores and dispersion for each group. If we find considerable variation in the average and dispersion of different groups, it indicates that our scale, although valid for some groups, is not generally applicable to as varied a group as that to which we have applied it. If this appears to be the case, we may either (*a*) abbreviate the original scale so as to retain only such indices as are found to have no cumulative variation for size of the unit measured, or (*b*) construct a separate scale for each population group within the limits of which the scale appears to be valid. Which of these alternatives to adopt must be governed by our specific purpose in each case.

6. As in the case of all other tests, no matter how carefully prepared, the final revision and validation of the sociometric scale must rest on an actual trial in the field. Such a test is important in several ways. First, it permits the elimination of possible overlapping of parts of the scale. If it is found that the score for several different factors of the units measured correlates consistently to a high degree in a large number of cases, then it is evident that scoring on any one of these factors alone is a satisfactory measure also of the other factors, to the degree that the correlation between them has been found to exist. Thus Chapin[9] found that a comparatively simple schedule for rating the socio-economic status of urban middle-class families could be constructed,

[9] F. S. Chapin, "A Quantitative Scale for Rating the Home and Social Environment of Middle Class Families in an Urban Community," *Jour. Educ. Psychol.*, 19:99-111, 1928. See especially the later revisions of this scale, the results of which are reported in F. S. Chapin, *Contemporary American Institutions,* (Harpers, 1935), Chap. 19.

and that the results correlated to a high degree with the scores obtained by several other more complicated schedules.

Second, the actual testing of the schedule in the field shows us which of the data called for are not available, are very difficult of access, or are too unreliable to warrant their use. The schedule must then be revised to eliminate these items, and, if possible, to include other more available data on the same aspects.

Finally, it is by actual use of the schedule in the field that the data necessary for the establishment of norms become available. We find in this way the average or modal score of a large number of organizations as well as the modal score of each aspect measured. The deviation of a particular case from this norm indicates its departure from the average, *i.e.*, from "reasonable" status or attainment. The dispersion in a distribution of scores of a large number of cases also enables us to determine the degree of reliability, *i.e.*, the probable error in measurements, made with our scale.

The best way to gain familiarity with the technique of construction of sociometric scales is to study the detailed steps involved in the construction of actual scales. Having considered at some length the principles governing the construction of sociometric scales, we shall now consider briefly some concrete examples of scales which have been developed to measure two different types of phenomena, namely, (1) health conditions and (2) socio-economic status.

F. THE MEASUREMENT OF HYGIENIC BEHAVIOR

The following exposition of the technique of constructing a scale for the measurement of hygiene in rural villages in Syria is admirable for its clarity, completeness, and attention to scientific standards: [10]

1. *Specifications*

a. Validity. A scale to be valid must measure in some demonstrable way what it claims to measure. This scale claims to measure (1) hygiene (2) of families (3) in Syrian (4) villages.

(1) To prove that the scale measures hygiene its scores must correlate with some index of health in a group of people, or com-

[10] S. C. Dodd, *A Controlled Experiment on Rural Hygiene in Syria,* (Oxford Univ. Press, 1934), pp. 2-12.

munities. High correlation coefficients in a population of defined range or variation are a necessary condition for such proof. But they are not a sufficient proof. For if two series of phenomena are causally related they must be correlated, but they may be correlated without any relation of cause and effect between them. In addition to the mutual association demonstrated by a correlation coefficient, there must be relation in time of antecedent and consequent, to show a cause and its effect. In interacting social phenomena it is difficult to separate two phenomena in a clear-cut test such as by making each act alone and see if the other follows. Fortunately, sociologists may build on the medical sciences, in which the causal relation of much of hygiene to health has been established experimentally. Therefore the agreement of medical experts that certain practices are "hygienic" ones and their correlation with health indices will be considered a necessary and sufficient proof of the validity of the scale.[11]

Next, the claim of the scale to measure *family* hygiene must be proved by showing that a score represents the practices of the family as a whole and is independent of the deviations of individual members. If the scores obtained from two different informants in each family correlate high it will be a proof of this claim.

The scale must be appropriate to conditions in *Syria* and similar regions of the Near East. For example, it must deal with anti-malarial practices, and not anti-yellow fever practices, as the former sickness is an important one here, while the latter is non-existent.

As conditions vary in different parts of the Near East, flexibility of the scale in order to adapt it to local conditions is essential. The answers to the questions must be recorded independently of any weighting in scoring in order to enable a different scoring or weighting system to be used in another locality where conditions may be different.

The scale must fit *rural* conditions. It must be possible to record and score all the variety of conditions, practices, and knowledge conducive to health which may be found in the villages. The distribution curves must not show a "pile up" of cases at either end due to the scoring system having a "cellar" or a "roof" which kept those cases from being spread further downwards or upwards. Validity in this respect will be proved if all

[11] "As some practices are of dubious or controversial value in medical opinion it is well, in case of doubt as to whether to include such items in the scale, to err on the side of over inclusion. It is easy to reject all such information in statistical treatment, but impossible to treat it if the original schedule did not secure it. Therefore the scale should include many items to make possible the later sorting of the more valid from the less valid. Minuteness in observing data is essential in science, although this minuteness goes far beyond the needs of social workers who may want to use such a schedule card later." (*Ibid.*, p. 2.)

answers given by Syrian villages can be recorded and scored on the scale without modes at either extreme of the frequency distribution graphs.

b. Reliability. To be reliable any measuring instrument must give the same results when reapplied to the same phenomena. Discrepancies in results are errors of measurement. If these errors tend to be in one direction they are constant, or one-way errors, but if sometimes in one direction and sometimes in the other, they are variable, two-ways, errors.

Constant errors are measured by differences between the average score of one observation, or survey, of a group and the average score of a second survey. This difference when divided by its standard deviation becomes the significance ratio. If this ratio is less than 3, scientific convention agrees that the observed error may be attributed to chance (sampling error). If the ratio is above 3, the probability of its occurrence by random fluctuations of sampling is approximately one in a thousand at most. . . This is so improbable as to mean with a high degree of probability that the observed difference is a "significant" constant error.

Variable errors may be measured by the correlation coefficients between the two surveys. A coefficient over .95 will here be considered to indicate high reliability or small variable error.

In order to eliminate the constant and the variable types of error, they may be analyzed into the following categories, classified by sources.

(1) *Sampling error.* A frequent source of error is from smallness of the sample. The formulas for the probable error, or for the standard error, of sampling measure the size of this error. To be reliable or "significant" all indices based on this scale must be larger than three times their standard error.

(2) *Seasonal error.* The smaller errors due to differences between informants, surveyors, etc., although included in the sampling error, may be analyzed out separately. It is better experimental technic to eliminate the component factors than to depend on the fact that sampling fluctuations tend to cancel each other when the sample is sufficiently enlarged.

As the scale seeks to measure the hygiene of a family throughout the year, it should not give different results at different seasons. For a reliable scale the results of surveys at different seasons must show a negligible constant error (as measured by the significance ratio) and a negligible variable error (as measured by the correlation coefficient).

(3) *Informant error.* The amount of informant error considered permissible is when the correlation between the scores derived from different informants in each family exceeds .90. It is possible to group the scores from men and women, or from adults

and minors, or from married and unmarried, separately, and thus determine the bias that sex, or age, or marriage, may introduce into the informant error.

(4) *Interviewer or surveyor error.* The permissible amount of error due to different surveyors interviewing a family may be specified as a correlation coefficient of not less than .90 between the scores from different surveyors. Where surveyors work in pairs this error may be split into a questioner error and a recorder error. The questioner error may be determined by repetition of the schedule on a group with a different questioner. The recorder error may be determined by having two or more recorders listen to the same informants and record the replies simultaneously but independently.

(5) *Scorer error.* The permissible amount of error due to different scorers may be specified as zero. Differences between different scorers can, and should be, completely eliminated by using the proper technics, hereinafter described.

(6) *Schedule error.* If a schedule putting the questions in different form were made up, after the manner of "alternate forms" of an intelligence test, the score of a family might become different. This schedule error, in its constant or variable form, is measurable by either the significance ratio or the correlation coefficient.

The sampling error does not always include the seasonal, informant, interviewer, scorer, and schedule errors. If in one set of scores there are included variations in season, informant, etc., the sampling error includes these. But if the set of scores keeps a contributing error constant, as in surveying at one season only, or with one interviewer, or with one scorer, or with one schedule only, then this error is not included in the probable error of sampling nor measured by it. This exposition suggests how these contributing errors may be eliminated and also how they may be measured separately or in combination.

A reliable scale should reduce all errors to a minimum, and then measure the amount of that minimal residual error. Only when the degree of accuracy of a scale is thus known can data yielded by it be of scientific value.

c. Objectivity and quantitativeness. Throughout the above specification of the reliability, (which has been set as a goal to be achieved by the constructor of this scale), there have been two principles implicit, but which need to be emphasized as corollaries of the specification of reliability. The first is that all the items must be cast in objective form with answers in definite printed categories to be checked such as "Yes" or "No," "1," "2," "3," *e.g.,* "Men ( ): Cereals ( ), Milk ( ), Vegetables ( ), Meat ( ), or Fats ( )," etc. If subjective and relative answers such as "Very much," "Good," etc., are to be inserted by the surveyor at his

own free will, there will be no standardization for the discovery of uniformities. Also differences between surveyors will increase the surveyor error enormously.

The second implied specification is that the data must be "quantifiable" — to coin a term expressing the idea that while the raw data may be secured in qualitative form it must be capable of being converted into quantitative form. Appropriate statistical treatment can then in scientific fashion induce patterns, interrelations, and uniformities with known limits of error. Exact study of the phenomena and simultaneous refinement of the instrument of observation then become possible.

d. Administration. Next in importance to the specifications of scientific validity and reliability of the scale are ease and simplicity of its administration.

(1) *Self-instructing.* The schedule that is put into the surveyor's hand should contain all necessary directions for its use and for ruling on unusual cases as they occur. These directions should accompany each item and should not involve consulting an index or manual of instructions or otherwise distracting the surveyor's attention from his conversation with his informant.

(2) *Lay surveyors.* The technic should be simple enough so that secondary school graduates may be trained in a week to carry it out. It should not require a doctor or trained nurse to administer.

(3) *Acceptability.* The questions should be acceptable to the informant, the government, the priest and other influential persons whose opposition, if aroused, would obstruct the gathering of data. To be acceptable to the informant questions of sex practices, income, ages in the zone of military conscription and some others must be avoided. To be acceptable to the government, questions of political import as to taxes, reactions to official steps towards sanitation or absence of them must be carefully phrased. To avoid other opposition, religious or other prejudices must not be unduly excited. Thus there should be no appearance of doubt as to the value to health and mortality of the religious pronouncements in regard to eating pork, circumcision, etc. To put this specification positively the schedule should win the cooperation of everyone through obviousness of its relation to health and the reduction of disease — a universal human desire.

(4) *Independence of language.* The schedule must be capable of being filled in from interviews in any language. It must not depend for its reliability on standardized phrasing of questions to such a degree as to become invalid when translated.

(5) *Inexpensive equipment.* Aside from the inevitable cost of transporting and maintaining a field personnel, the survey should keep costs of such equipment as is consumed, or used, down to a few piasters or a few cents per schedule filled in.

(6) *Brevity of interviews.* In order to increase the number of schedules filled in per day and so reduce the cost of the field personnel, the interviews should not require more than an hour apiece.

(7) *Convenience for statistical study.* The scale should be so printed and arranged as to facilitate recording, scoring, tabulating, correlating, etc., with a minimum of errors of eye and maximum of clerical speed. . .

2. *Technics of construction*

a. Acquisition of background of rural culture. To construct a scale with the above specifications, the author went to live for several periods of a week at a time in the villages of northern Syria. He also visited villages in Egypt, Palestine, Iraq, and Persia. He studied the Bedouins in the desert, and the Kurds in their tent encampments in the mountains. He visited families and drew them out in great detail on topics of health and sickness. His command of Turkish enabled him to have direct conversation in some Turkish speaking villages, while in the Arabic speaking villages he carried on most of the conversations through an interpreter. He accompanied the doctor and the nurse of an itinerant clinic on their visits, and in their systematic medical and physical examinations of every soul in two villages. He lectured, sometimes with the lantern, before the men and before the mothers on hygiene, leading them on to discuss their difficulties and points of view. He slept in their homes and in their tents and observed the habits of the family as to washing, defecation, sleeping, eating, and other routine patterns of life. He sat at the feet of the village barber, midwife, sheikh who dispenses charms, peddlers selling drugs, and occasional rural doctors, inducing them to instruct him out of their wisdom as to the details of plying their trades.

From these experiences and the reports of private and government clinics a list of the chief diseases was drawn up. Next, lists were prepared of the items, or patterns, of knowledge, practice, or environmental conditions, which seemed to be related to each disease or to health in general. In preparing this list it was necessary to avoid rejecting all folk practices as superstitious on the one hand, and to avoid accepting all western medical advice as sound in the given situation on the other hand. Folk remedies are occasionally [11a] shrewd cullings from the experience of generations, as in the use of pumpkin seeds for worms.

[11a] Striking historic examples of a kernel of truth in the chaff of folk superstitions are the discovery that fever is a remedy for incipient paretic paralysis from the practice by an African tribe of exposing such cases to nights in a malarial swamp ; and the discovery of digitalis as a heart drug from the folk reputation of a certain witch's brew of herbs.

Conventional hygienic advice of the west may backfire in a given situation. Thus taking baths in the village may be a positive menace to health. The villagers' idea of a bath is a steam heating process, the "Turkish bath," which requires facilities for slow cooling off. The mud hut has no such facilities and leaves the perspiring bather helpless in drafts. Experience seems to have given them the conviction of the woman who replied to the question "Why don't you bathe the baby?" by asking "Why should I kill it with pneumonia?" Hygienic slogans such as "Bathe often" need to be either discarded, or followed up with a new and detailed technic of bathing which adapts the rule to the local culture.

b. Trial of preliminary form of the schedule. From the lists thus compiled a trial list of questions was hectographed on a schedule card. After each question there were printed all the probable replies. Spaces were provided after every reply for checking those which might be encountered in a given family. These trial questions were taken to two villages and tried out. The doctor and the public health nurse in charge of a village clinic went over them carefully. A class in Methods of Social Research, composed of college seniors and graduates, was first instructed in the use of the schedule, then practiced it on some cases, and then turned in written criticisms.

c. Development of Form A and Manuals of Instruction. A revised schedule, labeled "Form A" emerged from these technics of criticism. It comprised two hundred and seventy questions. Most of these questions required subsidiary questions to elicit full information. About two thousand answers were printed to be checked affirmatively with a "v" or negatively with an "o." In addition, spaces were provided after questions calling for numerical answers. The questions were grouped in twelve sections, as follows:

Section 1. Identification data and vital statistics, including data on morbidity.

Section 2. Remedies for sicknesses represented by cuts, measles, typhoid, smallpox, colds, fevers, malaria, headaches, abdominal pains, diarrhea, constipation, childbirth, and venereal disease.

Section 3. Cleanliness — questions on the frequency and method of cleaning the person, the clothes, and the eating utensils.

Section 4. Sex — questions on sex practices and hygiene. These were not to be filled in routinely but only where confidential relations could be established as the questions were very intimate and touched on local, as well as universal, tabus.

Section 5. Food — an inventory of articles of diet within one year and the menu for twenty-four hours.

Section 6. Water — questions bearing on the source, amount, and purity of the family water supply in summer and in winter.

Section 7. Infants — a schedule on the infant's physical condition, diet, cleanliness, sleeping, and other habits.

Section 8. Property — an inspection of the rooms of the house, the yard and the stables.

Section 9. Waste — questions on the disposal of garbage and feces.

Section 10. Insects — inquiries as to the family's knowledge about, and practice in dealing with, mosquitoes, flies, lice and bedbugs.

Section 11. Sleep — questions on habits and facilities connected with sleeping.

Section 12. Heating — an inventory of equipment and practices for cooking and keeping warm in winter. . .

d. Construction of the scoring system. The answers on the schedule were recorded in terms of numbers, yes, no, and similar objective categories. To convert them into a single quantitative scale, credit points were assigned to each answer in a rough attempt to weight its hygienic importance, or relative contribution towards reducing the specific morbidity rates and general mortality rate in the local situation.

A thousand points were considered to provide ample discrimination. The worst hygienic conditions found were considered to set the zero end of the 1000 point scale. The best, or model, hygienic conditions which might be achieved with present knowledge and present economic standards of living in the country was set at the other extreme. Since unhygienic and even dangerous practices were given zero values, some positive credit was given to practices of neutral hygienic value. Thus the application of something generally aseptic, as urine, on a cut received less credit than the application of some antiseptic, but more credit than the application of something septic, such as scrapings of shoe leather. This in no way changed the relative credits given to hygienic, neutral, or unhygienic practices but only shifted the zero point of the arbitrary scale so that the scale dealt only with positive credits. This eliminated the more complicated algebra required in dealing with both positive and negative credits.

A panel of judges was then selected to allocate the number of points out of the thousand that should be assigned to each section of the scale. After full discussion they voted independently and the average number of points was found. These averages were slightly modified to round them off in order to make the total of the averages for the sections add up to 1000. Then each judge independently distributed the points, allocated to a section, over the individual questions of that section. Again the average allotment to each question was determined and modified to make the total come to the number of points assigned to that section. Then, the points allotted to each question were distributed over

the possible answers by a similar process. Provision was made for answers of equal hygienic merit which are alternatives (such as: the use of different vermicides; or the feeding of orange, tomato, or grape juice to babies, depending on availability).

Very often the points could not be assigned in a block — so many for a positive answer and nothing for a negative one — but were more suitably assigned in degrees. Thus for the seven points assigned to the use of soap (Question No. 62) the scoring formula was one point for every ten liters used annually. Again the hygienic value of certain acts depends on their frequency. Thus the five points for washing clothes were distributed: 5 for washing oftener than once a week, 4 for weekly washing, 3 for fortnightly, 2 for monthly, 1 for seasonally and 0 for never washing.

The allotment of points depended on subjective judgment. No more objective method seemed possible until statistical analyses, correlating the data with the criteria, should reveal weightings which corresponded more closely with objective reality. To reduce the variability of the judges' weightings (a) several *judgments* were averaged and (b) the *judges* were chosen for expertness in some aspect of the field of hygiene. . . All the judges participated in allocating points to the twelve sections and, (with one or two exceptions) to the questions within the sections. The first three judges then allocated the points within a question to its various answers.

The schedule prepared as described above was next pretested in ten villages and a city sample. The scores secured were graphed (*a*) for the whole scale, (*b*) the twelve section scores, and (*c*) for each of the 270 questions. These graphs were studied (*a*) to detect ambiguities in the questions; (*b*) to develop better categories in which to record the answers; and (*c*) to discover the range of answers so as "to provide for them adequately on the revised schedule without having a 'roof' or a 'cellar' which arbitrarily curtailed the distribution at one extreme or the other."[12] The graphs of six sample populations were further studied intensively in order to detect (*a*) defects of sampling, (*b*) spring-summer, (*c*) rural-urban and (*d*) Arab-Armenian differences, (*e*) the range of replies, and (*f*) unusual cases which might be suggestive as hypotheses. Each item was then considered separately from the standpoint of objectivity and its liability to any of the

[12] *Ibid.*, p. 20.

types of error considered above, for the purpose of selecting the items to be included in the final form of the scale. Finally, tests of the validity of the scale were carried out through the correlation of its scores with such indices as morbidity and mortality rates, and income. On the basis of all of the findings from this pretest, the entire scale was revised by adding some items, dropping others, and modifying still others. The result was a scale which conformed in a remarkable degree to the criteria set up at the outset.

G. A SCALE FOR THE APPRAISAL OF HEALTH WORK

A very different instrument from that described above but designed to measure somewhat the same general phenomenon in a radically different culture is the Appraisal Form for City Health Work developed by the American Public Health Association. The story of its origin has been summarized as follows:

> The desirability of some concrete standard of organized health work against which the work of communities should be compared has been appreciated for some time and several attempts have been made to construct a form that would serve as a health yardstick. More than ten years ago Dr. Charles V. Chapin, Health Officer of Providence, proposed a measure as the outcome of his study of state health departments. Later Mr. Franz Schneider of the Russell Sage Foundation prepared a table of relative values in health work. Dr. F. W. Sears of the New York State Department of Health developed a score card for rating city health work. These various standards embodied for the most part the personal opinions of individuals and of course there were many who held different opinions. To produce a document of some permanence there was needed something more than the judgment of an individual, even though that individual might be eminently fitted to express seasoned judgment.
>
> In 1920 the American Public Health Association through its Committee on Municipal Health Department Practice, began an intensive study of the work of city health departments. At the request of that committee, Dr. Chapin submitted a statement of standards and relative values in April, 1923. The committee had given much thought to these proposed standards and after amending the statement to harmonize with its collective judgment, approved it for experimental use by the Field Director, Dr. W. S. Rankin. Through contact with many city health officers, the Field Director incorporated many constructive sug-

gestions in the rating schedule as it appeared when it passed the committee.[12a]

The American Public Health Association and the American Child Health Association then got together and combined their respective plans. This statement of standards was then changed to meet the interests of the American Social Hygiene Association, the National Tuberculosis Association, the National Organization for Public Health Nursing, and the National Health Council. It was further modified in conference with three groups of city health officers from New England, Ohio, and Michigan, respectively. Final decision on each item was left to the vote of eight of these officers, representing a total of thirty-four in all.

The result was the first edition of the Appraisal Form. "After these lengthy preliminary conferences, after one full year of trial use by a group of nearly fifty cooperating city health officers, and after the application of the Appraisal Form to the study of the data for 186 cities collected by the United States Public Health Association surveys, a second general conference was held in New York in September, 1925, in which the Committee on Municipal Health Department Practice, the group of nine health officers representing cooperating officials in New England, Ohio, and Michigan, and the American Child Health Association, and other national voluntary health agencies, all took part. On the basis of recommendations made by this conference the committee has revised the Appraisal Form and believes that it is now sufficiently satisfactory to be issued with the approval of the committee as a basis for the appraisal of city health work."[13] The form thus revised was issued as the Second Edition of the Appraisal Form. A Third Edition, further revised, appeared in 1929. A special form for rural health work appeared in 1932. Finally the *Appraisal Form for Local Health Work* (first edition, 1938)[14] was developed in a form applicable to both rural and urban units. The accompanying illustration (Fig. 6) is taken from this latest edition.

[12a] *A Health Survey of 86 Cities,* American Child Health Association, Research Division, New York, 1925, pp. 15, 17.

[13] *Appraisal Form for City Health Work, Second Edition,* Introductory Statement, p. 2.

[14] American Public Health Association, 1790 Broadway, New York City.

The aim of the appraisal form is to provide "a brief schedule which will yield a reasonably accurate picture of the health services of a community as evidenced by certain typical sample activities. As originally developed, it was intended to cover primarily those services performed as organized functions of the health department or allied health agencies in a community. As it has developed, each revision has extended its original scope to a degree. The appraisal is deliberately not based on money expended or personnel employed, as these indicate resources rather than performance. Mortality rates have not been used directly as a base for rating since they are affected so frequently by racial, industrial, and other factors that they have no absolute value as standards of rating and their use makes comparison of communities frequently misleading. The idea in rating has been rather to measure actual accomplishment—as indicated by the extent to which: vaccinations are performed, infants are under medical supervision, physical defects of school children are discovered and corrected, cases of tuberculosis are hospitalized, vital statistics are properly recorded and analyzed, laboratory tests are performed—with confidence that the honest performance of such activities ultimately leads to the conservation of human life and efficiency, which is the goal of all public health work." [15]

The aspects of health work included in the Appraisal Form, and the relative weights assigned to each aspect are as follows:

	Weighting Factors
Vital statistics	40
Communicable disease control	160
Syphilis and gonorrhea control	90
Tuberculosis control	90
Maternity hygiene	90
Infant and preschool hygiene	170
School hygiene	140
Sanitation	90
Food and milk control	80
Bonus for balanced program	50
Total Weighted Score	1000

[15] *Appraisal Form for Local Health Work*, 1938, p. vii.

Each of these general divisions of health work are divided into from five to eight sub-categories such as (*a*) reporting and registration, (*b*) clinical service, (*c*) field nursing service, (*d*) institutional care, (*e*) laboratory service, (*f*) community health instruction, etc. After each item, there appears in the column "Value of item" a figure which indicates the relative importance of this item *in this part or section of the health program (e.g.,* tuberculosis control), as determined by group judgment. Each section is assigned a total of 100 points. The final score of any community is obtained by multiplying the actual score obtained in a given section by the proper weighting factor for that section as given above. The weighting factor expresses the relative value of that activity *(e.g.,* tuberculosis control) in the whole health program.

The accompanying extracts from the Appraisal Form will illustrate the method of scoring. It will be noted that standards are provided in each case. These standards do not represent ideal or Utopian norms, but reasonable, attainable standards under the present status of knowledge and development of public health work. The standards are not imaginary or theoretical, but are based on conditions actually existing at the present time in the large number of cities from which data for this schedule have been drawn. It will be noted that the data called for are always *objective* and *quantitative.* Finally, this schedule places the major emphasis on the *functional* aspects of the city health department — the actual work performed rather than the presence or absence of machinery for doing the work. For these and other reasons, the schedule is an excellent concrete illustration of the principles laid down earlier in this chapter regarding desirable characteristics of this type of sociometric scale and the methods of its construction.

The practical usefulness of an appraisal form of this kind has been admirably summarized by the American Child Health Association on the basis of the experience of a large number of cities:

As a brief for annual appropriations of the Department of Health. In several cities the rating schedule has been used to show in a comparative, objective manner certain deficiencies in the department of health, and the expenditures which would be

required to secure a full credit in place of these deficiencies. When presented in this manner to the appropriating body of the city, as a part of the health officer's budgetary request, the health department was, in one instance, the only department of the city government to receive an increase (amounting to 14 percent) for the coming year. In another instance, in the face of a strong movement for retrenchment throughout the city government, the health budget, presented in the light of the rating schedule, was the only departmental budget to remain uncut.

As a basis for a Health Program. The value of the rating schedule as an outline of a well-balanced health program has appealed to several health officers to such an extent that their future programs are definitely based upon achieving the standards set in the schedule. . .

As a basis for an Annual Report. It has occurred to one health officer to incorporate in the annual report of his health department the detailed rating of his city . . . and to publish the score attained with full discussion of many of the items. Another health officer has largely improved the character of his annual report in the light of the Appraisal Form.

As a means of interesting a Mayor or Chamber of Commerce. In a city which had recently scored itself, the health officer was studying the rating when the mayor of the city happened to enter the office of the health department. Upon being shown the low rating of the city in diphtheria prevention the mayor inquired what could be done to improve the situation. When the health officer explained that he could secure a full score (30 points) if he had $175 with which to buy toxin-antitoxin, the mayor promised to make the sum available from special funds. Other health officers have taken the rating of their city to the mayor or the chamber of commerce, and, by comparing the local rating with that of other cities, have awakened a new interest in the city's health record.

As a means of arousing and cementing the interests of the Departments of Health and Education and of the private agencies in each. A striking outgrowth of the use of the rating schedule in one city where the findings were reported in a special meeting, was the development of a new sense of the interdependence of all health agencies on each other and the formation of a joint health committee on which all agencies were represented under the leadership of the health officer. In another instance the rating schedule brought to the health officer for the first time the realization that he should cooperate with the department of education in its health work if his city was to secure the fullest return from its health activities. School medical inspection was promptly started.[16]

[16] Reprinted from *A Health Survey of* 86 *Cities,* (New York, 1925), p. 3. By permission of the American Child Health Association, New York City.

SURVEY SCHEDULE

SANITATION ACTIVITIES

I. GENERAL SANITATION, INCLUDING WATER SUPPLY AND EXCRETA DISPOSAL

The function of the health department in this field as in communicable disease control is rapidly swinging from one of law enforcement to one of education. It is only within recent years that the personnel of a health department responsible for the control of environmental sanitation has ceased to be known as "Sanitary police." They are still frequently referred to as sanitation inspectors and occasionally as sanitation officers. While state and local laws frequently provide the basis for the control program, recently developed sanitary codes covering these fields provide the necessary scientific and technical guide posts for a comprehensive educational program.

39. Inspection Service

The function of the sanitation personnel is primarily that of interpreting regulations and assisting individuals in complying with both the letter and the spirit of the law. While in some fields the service must assume responsibility for a certain amount of continuous supervision, in other fields the educational work of the department should decrease the necessity for routine supervision and the public should take more and more responsibility for maintenance of favorable sanitary conditions. On the other hand, when the health department through the establishment of inspectorial service in a certain field attracts public attention to its efforts in improving sanitary conditions, the public comes to rely upon this protection, and care must be exercised that this confidence in the department is not abused.

Number of sanitary inspections and reinspections (exclusive of smoke inspections, food inspections, and all communicable disease work)

	Number of sites	*Number of inspections*
a. Municipal water supplies		
b. Private premises		
Camp sites		
Swimming pools		
Barber shops		
Schools		
Plumbing[1]		
Sewage plants		
Other (specify)		
c. Malaria control		

d. Is there provision locally or through the state for the inspection of working conditions in industrial plants?

[1] This does not include routine plumbing inspections but only those of equipment and installation designed to insure against the possibility of cross-connections, actual or potential.

FIGURE 6. Sample pages from the Appraisal Form for Local Health Work. American Public Health Association, New York, First Edition, 1938. (*a*) Sanitation Activities.

APPRAISAL SCHEDULE

SANITATION ACTIVITIES

I. GENERAL SANITATION, INCLUDING WATER SUPPLY AND EXCRETA DISPOSAL

39. Inspection Service	Value of Item	Item for Scoring	Value Assigned HD	OOA	VA	Total
a. Inspections of municipal water supplies .	2					
Standard: 12 inspections per supply per year						
Inspections—12 plus . . . 2 *points*						
6–11.9 . . . 1						
Less than 6 . . . 0						
Inspections ..... ÷ number of municipal water supplies ..... = ... inspections per supply						
NOTE: In cities in which the responsibility for the protection of the water supply is fixed in a department other than the health department, this item may be satisfied if the health department takes samples for bacterial analysis at least monthly.						
b. Sanitary inspections	3	3				
Standard: 25 inspections per 1,000 population						
Inspections—25 plus . . . 3 *points*						
20–24.9 . . . 2						
15–19.9 . . . 1						
Less than 15 . . . 0						
Inspections ..... ÷ population ..... X1000 = ... inspections per 1,000 population						
c. Inspections for malaria control	2					
Standard: 50 inspections per 1,000 population						
Inspections—50 plus . . . 2 *points*						
30–49.9 . . . 1						
Less than 30 . . . 0						
Inspections ..... ÷ population ..... X1000 = ... inspections per 1,000 population						
NOTE: Where malaria shows an incidence of less than 10 cases per 1,000 population, this item should be omitted from consideration in appraising.						
d. Inspections of industrial plants	1*					
Standard: Provision locally or through the state for inspection of working conditions in industrial plants						

*For cities only.

FIGURE 6. (*Continued*). (*b*) Sanitation Activities.

SURVEY SCHEDULE

22. **Reporting and Registration**—continued

Number of cases and deaths from tuberculosis reported:

	19.....	19.....	19.....	*Average*
a. New cases reported before death [1] Pulmonary				
Other forms				
Cases reported after death Pulmonary				
Other forms				
Total cases reported				
Deaths, all forms				
b. Known living cases on register [2] Pulmonary				
Other forms				
Total				
Known childhood type cases on register				
Tuberculosis deaths in area occurring in sanatoria				

c. Is there regular and systematic reporting to the health department of all cases of tuberculosis returned to the area from state or other sanatoria? .....

d. Are epidemiological investigations and case histories made of all cases of tuberculosis reported which consider the case in relation to the family (household) and close associates in an effort to locate both source and spread contacts? .....

e. Number of contacts of new cases reported examined .....

[1] "Cases" here refers to those diagnosed as active, exclusive of childhood type.

[2] The register should be checked once a year to eliminate individuals who have moved from the area, died, or are reported clinically recovered from tuberculosis. Cases moving into the district and reported by health department of another area should be treated as new cases reported.

FIGURE 6, (*Continued*), (*c*) Tuberculosis Control.

APPRAISAL SCHEDULE

D. TUBERCULOSIS

22. Reporting and Registration	Value of Item	Item for Scoring	Value Assigned HD	OOA	VA	Total
a. New cases reported last year per annual average number of deaths for past 3 years	3	3				
Standard: 2 new cases per annual average number of deaths						
Cases — 2 plus . . . . 3 *points*						
1.5–1.9 2						
1.0–1.4 1						
Less than 1.0 0						
Cases reported ..... ÷ average number of deaths ..... = ... cases per death						
b. Known living cases last year per annual average number of deaths for past 3 years	3	3				
Standard: 5 cases per annual average number of deaths						
Cases — 5 plus . . . . 3 *points*						
3.5–4.9 2						
2.0–3.4 1						
Less than 2.0 0						
Known living cases ..... ÷ average number of deaths ..... = ... known cases per death						
c. Systematic reporting to health department of cases returned to area from state or other sanatoria	2	2				
d. Epidemiological investigation of all cases	3	3				
e. Contacts of new reported cases examined	4	4				
Standard: 3 contacts examined per new case reported						
Contacts — 3 plus . . . . 4 *points*						
2.5–2.9 3						
2.0–2.4 2						
1.5–1.9 1						
Less than 1.5 0						
Contacts examined ..... ÷ new cases last year ..... = ... contacts examined per case						

FIGURE 6. (*Continued*). (*d*) Tuberculosis Control.

H. THE MEASUREMENT OF SOCIO-ECONOMIC STATUS

Planes and standards of living have always been a subject of considerable interest in the social sciences. A large number of indices for the classification of families on this basis have been developed. The influence of home environment on personality and all manner of behavior traits has been a subject of protracted discussion. The need for an objective method of rating homes as desirable places in which to rear children first made itself felt in the field of social work. In this connection also a number of rating schemes and scales have been formulated. We select as illustrations several of the best of these instruments.

F. S. Chapin in 1928 defined socio-economic status as "the position an individual or a family occupied with reference to the prevailing average standards of cultural possessions, effective income, material possessions, and participation in the group activities of the community."[17] He constructed scales to measure each of these components. Cultural equipment was measured by assigning "arbitrary numerical weights to a list of articles such as books, newspapers, magazines, musical articles, and other items. The total of these weights constituted the culture score. Effective income was measured in terms of *ammains*."[18] Material possessions were measured by articles of household equipment weighted as in the first series. An index for participation in group activities was constructed on the basis of a detailed record of membership, contributions, attendance, committee service, and offices held in community organizations. Further analysis led to the conclusion that the total score given to living-room equipment might be taken as a fair index of socio-economic status, since it correlated highly with the combined score of the four indices; and therefore a scale consisting of 53 items of living-room equipment was constructed. The scale consisted of four groupings of items: Fixed Features, 14 items; Built-in Features, 5 items; Standard Furniture, 13 items; and,

[17] F. S. Chapin, "A Quantitative Scale for Rating the Home and Social Environment of Middle Class Families in an Urban Community," *Jour. Educ. Psych.*, 19:99-111, 1928.

[18] The *ammain* ("adult male maintenance") is a unit developed by E. Sydenstricker and W. I. King ("The Measurement of Relative Economic Status of Families," *Quart. Publ. Amer. Statis. Asso.*, 17:842-57, 1921).

Furnishings and Cultural Resources, 21 items. On the basis of experience, group two (Built-in Features) was eliminated; and later, in 1931, a group of items designed to reveal the conditions of the room and its contents was added." [19]

After considerable experimentation and use of the scale in various parts of the country, Chapin decided that the scale could be greatly simplified. The third revision, *The Social Status Scale* 1933,[20] reproduced herewith was the result. Of the original 53 items, 17 were selected because of their demonstrated capacity for differentiating between social classes. They were assigned new weights so as to give approximately the same total as was given by the original 53 items. The resulting scale showed a reliability of $r = .96$ on a sample of 50 professional men's homes by the test-retest method, and a validity measured by correlation ratios of .57 with occupation and .44 with income, for 442 homes. Numerous other tests of the validity of this scale have been applied with good results. The following figures, based on representative samples, are the standardized mean scores on the 1933 scale for different types of homes: [21]

	Mean Score
Professional men's homes (white) in Minneapolis	147
Boarding homes in Minneapolis (white) approved by the Minnesota State Children's Bureau	126
Miscellaneous average homes (white) in Indianapolis	123
Upper-class Negro homes in Minneapolis	107
Industrial-class homes (white) in urban Alabama	55
Industrial-class homes (white) in urban Massachusetts	50
Poor relief cases (white) in Minneapolis	47
Poor relief cases (Negro) in Minneapolis	40

The present author also tested the validity of the Chapin scale by comparing the score distribution of the scale for the population of a small village with the common-sense judg-

[19] Wm. H. Sewell, *The Construction and Standardization of a Scale for the Measurement of the Socio-Economic Status of Oklahoma Farm Families*, (Technical Bulletin No. 9, Oklahoma Agricultural and Mechanical College Agricultural Experiment Station, Stillwater, Okla., April, 1940), p. 15.

[20] F. S. Chapin, *The Measurement of Social Status by the Use of the Social Status Scale* 1933. (Univ. of Minn. Press, 1933. 16 pages, 25 cents.) See also, F. S. Chapin, *Contemporary American Institutions*, (Harper, 1935), Chap. 19.

[21] For a full discussion of the validity of this scale, see F. S. Chapin, *Ibid.*, Chap. 19. A modification of the weights for the different items in the scale of 1933 on the basis of a factor analysis is now under way.

THE SOCIAL STATUS SCALE 1933

REVISED, 1936

VISITOR'S NAME.....

SCHEDULE NUMBER..... DATE.....

FAMILY NAME.....

STREET ADDRESS.....

TOWN AND STATE.....

RACE.....

DIRECTIONS TO VISITOR

1. The following list of items is for the guidance of the recorder. Not all the features listed may be found in any one home. Entries on the schedules should, however, follow the order and numbering indicated. Weights appear (in parentheses) after the names of the items. Disregard these weights in recording. Only when the list is finally checked should the individual items be multiplied by these weights and the sum of the weighted scores be computed, and then only after leaving the home. All information is confidential.
2. Where the family has no real living room, but uses the room at night as a bedroom, or during the day as a kitchen or as a dining room, or as both, in addition to using the room as the chief gathering place of the family, please note this fact clearly and describe for what purposes the room is used. Scores corrected for these uses are as follows: (1) used as dining room, deduct 6; (2) used as kitchen, deduct 9; (3) used as bedroom *or* dining room and kitchen combined, deduct 12; (4) used as bedroom, dining room, and kitchen combined, deduct 15.
3. If the item you are scoring is present, enter the figure 1 on the *dotted* part of the line. If more than one item is present, write 2, 3, 4, etc., as the case may be. If the item is not present, enter a zero. Each division of each of the twenty items should have a number (0, 1, 2, 3, 4, etc.) entered opposite it. *Be sure that each item gets some number.* Leave no dotted line blank. A blank line means an oversight on your part.
4. Disregard figures in parentheses until all schedules are finished.
5. Always enter color of each family (white or Negro; specify if any other) below town and state.
6. Descriptions of items in Parts I and II.
   (1) Hardwood floors are usually made of narrow boards. Verify your own judgment by asking the occupant.
   (2) A large rug is one covering the entire floor area except a border of about 12 inches. (Usually a 9 x 12-foot rug.)
   (3) A drape is a covering over a window or windows (with or without curtains). The drape may be either at the side of the window or across the top. Each draped window gets one point. Where the drape covers more than one window, give the question as many points as there are windows covered by the drape.
   (4) Only a real fireplace (one in which a coal, wood, or other fire, including gas, can be built) gets a number.
   (6) A library table is any table not used for serving meals. (Exclude such small tables as end tables, card tables, etc.) If a library table is used as a personal-social desk, number it either as a library table or as a personal desk, *but not as both.*
   (7) Armchairs include rocking chairs with arms.
   (8) Piano bench: a chair or stool does not get a point.
   (9) See No. 6. A personal-social desk is a writing desk.
   (10) If no bookcase is seen, be sure to ask if there is one in another room. Be sure to record any bookcase, whether in living room or in any other room, except professional library of doctor, lawyer, clergyman, teacher.
   (11) Sewing machine must be in living room in order to be marked minus two points. If outside living room, mark zero.
   (13) Alarm clock: same as No. 11.
   (14) Number of different periodicals *regularly bought* or *subscribed* to. Always get answer to this question by asking.
   (15) Newspapers: same as No. 14.
   (16) Telephone: to be recorded whether in living room or in some other room. Always ask concerning the telephone if you do not see one.
   (17) Radio: to be recorded whether in living room or in some other room. Always ask regarding the radio.
   (18) It is possible for a room to be both (a) spotted or stained and (b) dusty. In this case each gets *one* check.
   (20) "In good repair" means in good condition and not obviously patched up; therefore only one of the parts (a, b, or c) gets a check.
   (21) Only one of the parts (a, b, or c) gets a check.

Reprinted, with revisions, from *The Measurement of Social Status* by F. Stuart Chapin, The University of Minnesota Press, Minneapolis, 1933.

(205)

FIGURE 7. The Social Status Scale 1933. (Revised, 1936.)

PART I. MATERIAL EQUIPMENT AND CULTURAL EXPRESSION OF THE LIVING ROOM OF THE HOME

1. Floor, softwood (6) _____
   hardwood (10) _____
2. Large rug (8) _____
3. Windows with drapes (each window 2) _____
4. Fireplace with 3 or more utensils (8) _____
5. Artificial light, electric (8) _____
   kerosene (−2) _____
6. Library table (8) _____
7. Armchairs (8 each) _____
8. Piano bench (4) _____
9. Desk: personal-social (8) _____
10. Bookcases with books (8 each) _____
11. Sewing machine (−2) _____
12. Couch pillows (2 each) _____
13. Alarm clock (−2) _____
14. Periodicals (8 each) _____
15. Newspapers (8 each) _____
16. Telephone (8) _____
17. Radio (8) _____

Score on Part I _____

PART II: CONDITION OF ARTICLES IN LIVING ROOM

To provide some objective rating of qualitative attributes of the living room, such as "aesthetic atmosphere" or "general impression," the following additional items may be noted. The visitor should check the words that seem to describe the situation. Some of the weights are of minus sign, and so operate as penalties to reduce the total score of the home.

18. Cleanliness of room and furnishings
    a. Spotted or stained (−4) _____
    b. Dusty (−2) _____
    c. Spotless and dustless (+2) _____
19. Orderliness of room and furnishings
    a. Articles strewn about in disorder (−2) _____
    b. Articles in place or in usable order (+2) _____
20. Condition of repair of articles and furnishings
    a. Broken, scratched, frayed, ripped, or torn (−4) _____
    b. Articles or furnishings patched up (−2) _____
    c. Articles or furnishings in good repair and well kept (+2) _____
21. Record your general impression of good taste
    a. Bizarre, clashing, inharmonious, or offensive (−4) _____
    b. Drab, monotonous, neutral, inoffensive (−2) _____
    c. Attractive in a positive way, harmonious, quiet and restful (+2) _____

Score on Part II _____ Total score,* Parts I and II _____

* With penalties deducted. (See paragraph 2 on the other side.)

FIGURE 7. (*Continued*).

ments of a banker and a janitor both of whom had always lived in the village. Figure 8 indicates the degree of correspondence between the scale scores and the subjective judgments.[22]

Chapin's *Social Status Scale* 1933, is outstanding as a brief, simple, reliable, and valid instrument for measuring socio-economic status of urban families.[23] Its brevity and emphasis on material possessions increases, of course, the possibility of erroneous or less discriminating scores in unusual, individual families. The remarkable fact is that, in spite of the above limitations of the scale and the further fact that it was originally constructed from comprehensive data from a sample of only 38 urban middle-class families in Minneapolis, it has shown remarkable reliability and validity in several thousand cases from a dozen widely scattered states. These results suggest (*a*) that there may be rather *simple and stable elements* in social situations of great apparent complexity (*e.g.*, socio-economic status); and (*b*) that *the discovery of*

[22] G. A. Lundberg, "The Measurement of Socio-economic Status," *Amer. Sociol. Rev.*, 5:29-39, 1940. For purposes of this test the Chapin scores were translated into a six point scale according to the following equivalents:

Upper Class	1. Upper part — 250 and up
	2. Lower part — 200-249
Middle Class	3. Upper part — 150-199
	4. Lower part — 100-149
Lower Class	5. Upper part — 50-99
	6. Lower part — 0-49

The reasons for disagreements in individual cases were also analyzed. (*Ibid.*) The correlation coefficients (uncorrected eta) were as follows: Chapin score and janitor's rating, $\eta = .51 \pm .04$; Chapin score and banker's rating, $\eta = .61 \pm .03$.

TABLE V

Frequency Distribution of Ratings of the Socio-Economic Status of 196 Families on a Six-Point Scale by (a) *The Chapin Scale*, (b) *A Local Janitor, and* (c) *A Local Banker*

Rated by	*Number of Families Assigned to each Gradation of the Six-Point Scale*						*Total*
	1	2	3	4	5	6	
The Chapin Scale	0	10	72	68	42	4	196
The Janitor	3	10	42	81	50	10	196
The Banker	22	15	88	49	22	0	196

[23] For a digest of the results of the use of this scale in various parts of the United States, see F. S. Chapin, *op. cit.*, ch. 19.

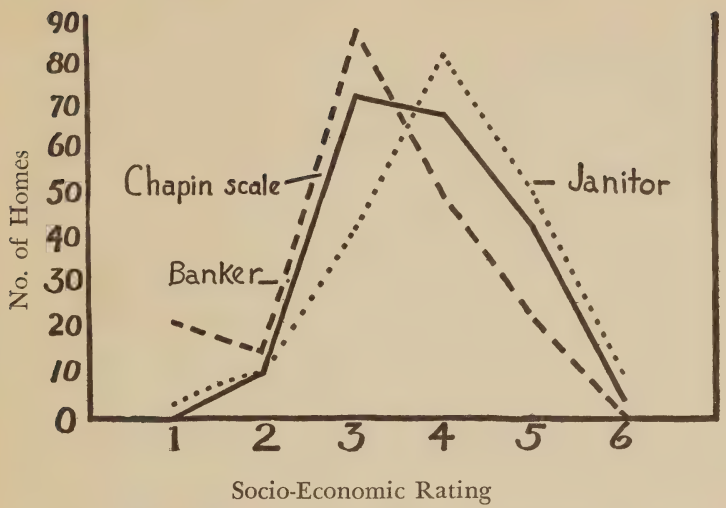


FIGURE 8. Comparative Socio-Economic Rating of 196 Homes by a Janitor, a Banker and the Chapin Scale.

these elements and their reduction to indices is the key to practical social measurement.

Another excellent scale directed at somewhat the same problem as the Chapin scale but including more items and a broader variety of aspects is the *Minnesota Home Status Index* published by Alice M. Leahy in 1936.[24] The scale was based on 600 urban white families with children in the age group 5-14. The families were selected from various occupational groups, and were scored on a total of 84 items. The differentiating power of each item was tested by the criterion of internal consistency. The fifty items finally selected were assigned weights by the sigma method. (These techniques will be discussed below.) The final scale consists of the following categories: Children's Facilities, 11 items; Economic Status, 13 items; Cultural Status, 11 items; Sociality, 13 items; Occupational Status, 1 item; and Educational Status, 6 items. The scale showed a reliability of +0.92 (corrected

[24] *The Measurement of Urban Home Environment,* (Univ. of Minnesota Press, 1936).

by the Spearman-Brown formula) by the method of split halves. Its validity is reflected by a correlation of +0.94 with the Sims Score Card [25] for 200 homes, and the wide variation in scores shown by the scale when applied to widely separated economic groups. *The Home Status Index* is one of the most carefully constructed and statistically adequate of existing scales.[26]

These scales were designed for urban families. A similar instrument to measure the socio-economic status of farm families was developed by William H. Sewell and published in 1940.[27] Data on 123 items from 800 farm families in three counties of Oklahoma served as a basis for the construction of the scale. Sewell utilized all the experience of previous attempts and the construction of such scales as those reviewed above as well as several decades of earlier experience, and his methods must be regarded as a model of how to construct a scale of this kind. The student is urged to study the full text of Sewell's original account, as only a summary of his procedure is possible here.

I. THE PRINCIPAL STEPS IN THE CONSTRUCTION OF SOCIOMETRIC SCALES

1. *Item validation.*[28] The first problem in the construction of a sociometric scale is to select items that will differentiate various degrees of the phenomenon measured. As we pointed out earlier in the chapter, the original selection of items (really in the nature of hypotheses) can best be done by people thoroughly familiar with the field to be measured. When a tentative list of items has been selected, their discriminatory value for purposes of measurement must be

[25] V. M. Sims, *The Measurement of Socioeconomic Status,* (Public School Publishing Co., Bloomington, Ill., 1928).

[26] Two other scales aiming at somewhat the same objectives should be mentioned: M. J. McCormick, *A Scale for Measuring Social Adequacy,* Social Science Monographs, I, No. 3, Washington, D. C., 1930; E. M. Burdick, "A Group Test of Home Environment," *Archives of Psychol.,* No. 101, 1928.

[27] Wm. H. Sewell, *op. cit.*

[28] For a review of this subject and an evaluation of the various techniques, see J. H. Long, P. Sandiford, *et al., The Validation of Test Items,* (Educational Research Bull. No. 3, Univ. of Toronto Press, 1935).

tested. Three general methods of testing items for this purpose are in common use.[29] One consists of correlating test items with some acceptable external criterion (judgments of experts, other tests, etc.) Another method consists of correlating individual items (usually by the "bi-serial *r*" method) with the test as a whole or with particular internal criteria. A third method, which avoids the laboriousness of the correlation techniques and gives substantially the same results, is the "critical ratio" method. This consists of a comparison of the frequency of occurrence or non-occurrence of each item in relation to the total score in selected segments of the distribution. The procedure is as follows: (1) A total score is calculated for each person taking the test, (or for each family scored). (2) Papers are placed in rank order on the basis of total score. (3) Equal segments at the extremes of the distribution are chosen. (4) Percentage differences between these segments are then calculated for each item. (5) The statistical significance of the difference is determined for each item (critical ratio). (6) Items with largest significant differences are said to have great discriminating power and are retained for the final scale.[30] This method assumes that the items are normally distributed in the general population, and, (as is true of all criteria of internal consistency) that the total score is an adequate criterion for validation.

Some limitations of tests of internal consistency as measures of validity of test items have been noted by Sletto[31] and Mosier:[32] (1) When a small sample is used chance errors may play a large part in the ranking of items in discriminative value from group to group. (2) If the sample and the number of items in the preliminary scale are small, a con-

[29] A third method for selecting test items, namely, factor analysis, should be mentioned as of possibly great importance. Its principal practical difficulty at present is its laboriousness in cases where the number of items and the sample are large. See L. L. Thurstone, *The Theory of Multiple Factors*, (Edwards Bros., Inc., Ann Arbor, Mich., 1933); *Simplified Multiple Factor Method*, (Univ. of Chicago Book Store, 1933); *The Vectors of Mind*, (Univ. of Chicago Press, 1935).

[30] *Cf.* Sewell, *op. cit.*, p. 31. The method is described further below.

[31] R. F. Sletto, *The Construction of Personality Scales by the Criterion of Internal Consistency*, (The Sociological Press, Hanover, N. H., 1937), pp. 81-83.

[32] C. I. Mosier, "A Note on Item Analysis and the Criterion of Internal Consistency," *Psychometrica*, 1:275-82, 1936.

siderable error in the discriminative value of an item may arise from including the item in the score on the preliminary scale when determining its discriminative value. (3) Statistically significant critical ratios for all items in a scale do not assure that all items are measuring a single common factor.[33]

"However, these limitations need not invalidate the use of the criterion of internal consistency in test construction. The first weakness may be overcome by the use of a large sample. Sletto has shown that chance errors play a small part in the discriminative value of items when the sample includes as many as 400 cases. The second limitation may be overcome by the use of a technique devised by Sletto to compensate for the displacement increment. Since the displacement increment bears an inverse relation to the number of items included in the preliminary scale, it may also be minimized by using a large number of items. The third limitation is probably inherent in all techniques other than that of factor analysis. The only way known at present to make sure that each item in a scale is measuring a single common variable is to make use of the complicated factor analysis technique." [34]

By the above methods it is possible to select those items which discriminate most sharply between different quartiles (or other divisions) of the sample from which data for the construction of the scale is drawn. For example, if, as Sewell found, the item "separate living room" is present in only 5.5 percent of the cases in the lowest one-fourth of the sample (as measured by the total score), in 25 percent of the next lowest one-fourth, in 48 percent of the third quarter, and in 82.5 percent of the highest scoring quarter of the sample, this indicates that the particular item has high and consistent discriminative value. Of course, many items showing less extreme differences may still be of discriminative value. The statistical significance (*i.e.*, the probability that a difference is due to other than chance fluctuation in the sample) of the difference between two percentages may be computed

[33] *Cf.*, R. F. Sletto, *op. cit.*
[34] William Sewell, *op. cit.*, p. 31.

by simple methods described in any standard statistical text.[35] When the probability that a given difference between two percentages might be due to chance is only 1 to 5 in a 100 (*i.e.*, when the critical ratio is from 3 to 2), we assume that the observed percentage difference is highly likely to reflect a difference of the kind we wish our items to indicate. We would consequently include such items in our scale. When this criterion was applied to *all percentage differences in the presence or absence of items in successive quartiles* in Sewell's scale, 82 items of the original 123 were found not to be discriminatory to the required degree and hence were eliminated.[36]

2. *The weighting of items.* When the items of a scale have been selected, scores or weights must be assigned to them. The simplest method that suggests itself is to assign zero to the absence of an item and 1 to its presence. In the case of graduated items, multiples of 1 may be assigned for successive levels. This assignment of values is, of course, entirely arbitrary, but in many cases yields as satisfactory results as more complicated and theoretically defensible methods.

Leahy[37] describes a second method, the difference method, based on the reliability of the percentage differences of possession among successive quintiles of the population on which the scale is based. This is in part the same procedure as has been described above in the determination of the discrimi-

[35] When it is desired to determine the significance of a large number of differences, as in the construction of a test, the Edgerton-Patterson tables should be used. H. A. Edgerton and D. G. Patterson, "Table of Standard Errors and Probable Errors of Percentages for Varying Numbers of Cases," *Jour. Applied Psych.*, 10:378-91, 1926. (Obtainable from the Editor of the above journal for 25 cents.)

The critical ratio is computed as follows:

(1) The standard error of the difference of two percentages

$$\sigma \text{difference} = \sqrt{\sigma\rho_1^2 + \sigma\rho_2^2}$$

(2) The critical ratio

$$= \frac{\text{difference between the two percentages}}{\text{standard error of the difference as in (1)}}$$

$$= \frac{\text{difference}}{\sigma \text{ diff.}}$$

[36] Five others were dropped for practical reasons leaving a total of 36 items in the final scale divided as follows: I. Material possessions, 15; II, Cultural possessions, 13; III. Social participation, 8.

[37] Alice M. Leahy, *The Measurement of Urban Home Environment*, (Univ. of Minnesota Press, 1936), pp. 41-42.

natory power of different items. The weight assigned to an item is the average of the critical ratios secured from a determination of the significance of the percentage differences between successive portions (in this case quintiles) of the population.[38]

A third method, and that finally adopted by both Leahy and Sewell, is the sigma method. This method weights an item in inverse relation to its frequency in the total population. Items that occur rarely are given the greatest weight, and absence of an item, as well as presence of it, receives a score value. Leahy states the assumptions of the method as follows: "As in the difference method, each item is assumed to be normally distributed. *In addition, when, as in this study, the possession of an item is regarded as desirable, there is the assumption that possession deviates on the positive side of the mean of the whole distribution with* 100 *per cent as its termination point.* On the other hand nonpossession of the item deviates in a negative direction from the mean of the whole distribution with the 50th percentile as its termination point. Furthermore, it posits that the most typical figure for percentage of either possession or nonpossession is one-half the observed percentage frequency."[39]

Sewell gives the following illustration of the application of this method:

"The calculation of the sigma score may be illustrated by computing the score for an item of the scale, such as item 38 'radio.' This was possessed by 46.1 percent and not possessed by 53.9 percent of the sample families. The calculation of the score for possession of the item follows:

$$100 - \left\{ \frac{46.1}{2} \right\} = 76.9$$

The distance of 76.9 percent, which is +0.737 sigmas, is read from a table of values of the normal probability integral.[40]

[38] $\dfrac{\Sigma \dfrac{D}{\sigma \text{ diff.}}}{5}$

[39] Alice M. Leahy, *op. cit.*, p. 42.

[40] The table presented by E. L. Thorndike was used in this study. See *An Introduction to the Theory of Mental and Social Measurements*, (2nd edition, Columbia Univ. Press, 1913), Table 44, p. 198. However, any standard table for areas under the normal curve may be used.

This may be used as the score for the possession of a radio. For non-possession the score is computed as follows: the typical value of the observed frequency is determined, in this instance $53.9 \div 2 = 26.9$; then, regarding the 50th percentile as the termination point for this segment of the distribution, the sigma value of 26.9 percent is read from the table. This is —0.615 sigmas and may be regarded as the score for non-possession of the item.

"In the case of graduated items, the cumulative percentage frequency distribution of this item is determined and the end of each successive truncated section is considered as the termination point from which the typical value would be located. For example, item 60, 'magazines,' has the following percentage distribution and sigma values:

Number of Magazines Taken Regularly	*Number of Families*	*Percentage Frequency*	*Cumulative Percentage Frequency*	*Sigma Values*	*Score Values*
6 and up	26	3.3	100.0	+2.144	8
4–5	96	12.0	96.7	+1.323	7
2–3	276	34.6	84.7	+0.451	5
0–1	400	50.1	50.1	−0.675	3

"As is apparent from the examples, sigma scoring produces negative as well as positive weights. However, all scores may be made positive by adding a constant to each of the original sigma weights. To increase further the magnitude of the resulting scores, all may be multiplied by a constant. In the present case, 2 was used as the constant in both operations. When this procedure is applied in the above examples, the score for possession of a radio becomes 5 and the score for non-possession becomes 2. The scores for possession of varying quantities of the graduated item, 'magazines,' are shown in the extreme right-hand column of the example above."[41]

In conclusion it should be said that the principal advantage of the sigma method is that it expresses scores in standard units. Both Leahy and Sewell found, however, that the sigma method yielded results almost identical with either the

[41] Wm. Sewell, *op. cit.*, pp. 43-46.

simple scoring or the difference methods described above.[42]

3. *Standardization of a scale.* When a scale shows satisfactory reliability and validity we say that it is standardized. Reliability refers to the ability of a scale to measure *consistently, i.e.,* repeatedly and in the hands of different investigators, whatever it purports to measure. Validity is usually interpreted to mean simply that a scale measures that which it professes to measure, or which we want it to measure. We have dealt with both concepts, and the conventional tests of each, in several previous connections. The tests of reliability are relatively simple, objective, and non-controversial. The question of validity, however, involves some theoretical questions of utmost importance to the future of the kind of measurement discussed in this book. We conclude, therefore, with a brief consideration of this subject.

What people usually mean by validity is that a scale purporting to measure something (designated by a familiar word) should agree with common-sense judgments of the phenomenon. That is, we expect a scale of socio-economic status to yield low scores for the beggar and for people on relief, and high scores for the banker and the factory owner. There can be no doubt about the general usefulness of this ever-ready, common-sense test of validity. In the absence of more formal and crucial tests, common-sense estimates of the degree to which an instrument aids the adjustment sought in concrete cases will always be a practical test of the validity of the instrument. But this fact should not deceive us into assuming that the validity of scientific instruments depends always or entirely upon their conformity with the immediate findings of common-sense. Indeed, this is a criterion very generally disregarded in fields where well-recognized measuring instruments have been developed. For example, no one would think of regarding his own common-sense impressions of temperature, weight, distance, or composition of the liquid in the radiator of his car, as more reliable than the readings of the appropriate instrument in each case. What, then, have been the criteria upon which the validity of these instruments have gained acceptance?

[42] *Ibid.,* p. 43, footnote 22. The correlation coefficients found between the results of the different methods by both investigators were all +0.98 or above.

The crucial test has been whether the readings of the instruments, when applied to the solution of some adjustment problem other than the mere confirmation of common-sense, have contributed toward the solution of this larger problem, toward the solution of which the individual measurement in question was only a means. That is, the criterion of the validity of the hydrometer is not whether it confirms our common-sense estimate based on smell or taste as to the amount of alcohol in the solution in the radiator of our car. The criterion is whether an acceptance of the reading of the instrument and acting on it keeps the radiator from freezing in a higher proportion of cases than by following our common-sense impressions. The same point could be illustrated even more strikingly with reference to the measuring instruments of height and weight. Note that the conformity of the readings of the hydrometer to our common-sense impressions of the amount of antifreeze liquid in the radiator is in this case wisely regarded as *completely irrelevant* as a test of the validity of the instrument.

The assumption that instruments for the measurement of social phenomena to be valid must necessarily conform in their results to a consensus of common-sense findings is, therefore, entirely unwarranted. This does not mean that when, as in the case of the Chapin scale which has been described, the results of common-sense estimates and standardized scales are in high agreement, this may not be of interest and value. We are merely emphasizing that this criterion is not a necessary or even a primary one. It is not a complete or final test of the validity or usefulness of measuring instruments of any kind. The more important criterion is the capacity of the instrument to achieve with high reliability *any significant designated type of result.* The conformity of the readings of standardized scales to our common-sense reactions to the (presumably) same phenomenon may be said to be chiefly of esthetic or linguistic rather than of scientific interest.

From the beginning of psychological, "mental," and social measurement, the argument has persisted in some quarters that, however clever the instruments, and consistent and practically valuable the results of their use, there remains, nevertheless, the theoretical obstacle that we do not know "what

they measure." The feeling seems to be that this "what" must be some entity outside and beyond *the conditions and behaviors (the stimuli) referred to by the symbols of the scale.* This is an illusory problem springing from untenable philosophical preconceptions, and more especially from a misunderstanding of the nature of all symbolic, linguistic, and semantic phenomena. But the view is sufficiently common to warrant our attention.

When a new measuring instrument is invented in any field, the question always arises, "What does it measure?" Since a measuring scale is a symbolic device consisting of a serial set and system of symbols for the designation of a gradation of certain kinds of conditions, qualities, events, or behaviors, the question of what is referred to by the scale is, of course, not fundamentally different from a similar question with regard to any new word which may be proposed to designate some phenomenon. The answer in the latter case, on elementary levels, is to point, gesticulate, or engage in some other overt and non-linguistic form of behavior. That is, as we shall see, also the only final way to answer the question of what a scale measures. In the case of folk-language designations and in the case of well-established measuring devices, however, we learn these meanings as we learn language in general and never have occasion to analyze exactly what we mean by them. As a result, we frequently reify these designations into entities, which are then assumed to have a self-sufficient existence quite apart from the conditions, qualities, events, and behaviors which the word was originally invented to designate. The word "intelligence," as generally used today, is a good illustration.[43] Any attempt to formalize, objectify, and render more explicit such a term is, therefore, likely to encounter the objection that the referent (*e.g.,* behaviors or conditions considered in constructing

[43] See, for example, M. May, "Ten Tests of Measurement," *Educ. Rec.,* April, 1939. "Suppose the candidate is intelligent. How can we be sure that it exists (*sic*) as a detectable quality of human beings in the sense that mass is a detectable quality of objects in the physical world? . . . From the standpoint of common sense, this is a foolish question. Everybody, in civilized cultures at least, knows that it does exist . . . (p. 205). . . How can the psychologist be sure that it [the intelligence test] detects only intelligence and nothing else?" (p. 207). I have discussed this subject at some length in *Foundations of Sociology,* (Macmillan, 1939), pp. 446-57.

a scale) designated by the new term unfortunately *is not* the original entity. *Of course,* it is not. Symbols never *are* what they designate. A slightly more sophisticated objection is that the new symbol does not refer to all referential aspects or implications of some well-known term which the new term aims to supplant. If the new term has been adequately defined, it will make explicit exactly what it does or does not cover. We do not thereby declare omitted aspects unimportant or less significant. The omitted aspects or referents may be designated by some equally explicit additional symbol.

Consider, for example, how readily the term "socio-economic status" passes from mouth to mouth even among sociologists and social workers on the assumption that all competent people know what it designates. When a formal and rigorously defined scale for measuring socio-economic status is invented and applied, the question arises as to "what," after all, it measures; whether what it measures is *really* socio-economic status, etc. Why do these appear to be difficult and important questions in the case of the scale, but more or less self-evident in the case of our everyday use of such a term as socio-economic status? The question of what is designated by the term is a highly proper one *in both cases.* The amusing thing is that only in the case of the scale can an adequate answer be given, whereas the reverse is commonly implied to be the case. Scales had their origin in precisely this fact that the more perspicacious individuals realized that it is only in such form that we can answer the question of *what* we mean by given expressions of gradations in socio-economic status. This meaning, *in terms of the conditions and behaviors taken into consideration by the scale,* is also the only possible answer to the question of what any scale or any informal judgment measures.[44] The statement

[44] Possibly the reason for misunderstanding on this subject is failure to keep in mind that all the operations in the construction and calibration of a measuring device *are part of the measurement process.* Every time we use a measuring instrument, the operations involved in its calibration are implicitly present. Another possible reason for the misunderstandings regarding such a statement as "intelligence is what the tests test" may be the assumption that *only a certain way of doing the particular acts* involved in taking a given test (*e.g.,* putting round pegs in round holes) constitutes the referent of the word "intelligence." Actually, the word is used to designate a way of doing these particular acts *and* certain ways of doing *all other acts*

that socio-economic status *is* what a scale for measuring socio-economic status measures, has therefore the same validity as to say that the conditions and the behaviors which any group calls high socio-economic status *is* high socio-economic status for that group. To deny either statement and to argue or imply that, in addition to the above explicit and avowedly relativistic definition, there is an absolute definition of socio-economic status (what status "really *is*"), independent of any culture, conditions, or behavior, is surely a linguistic delusion. It is like arguing that we are making a mistake in calling water wet, because it is really not wet at all. Formal instruments of measurement, by explicitly recognizing the conditions, behaviors, and standards involved in all measurement, strip the mystery from many words and hence cause temporary resistance to the acceptance of new instruments. To assume that any scales whatever can be said to measure some more ultimate "what" than the conditions, events, and behaviors its symbolism refers to, is pure mysticism, not to say a type of superstition.

A further reason for the above confusion seems to be that formal scales undoubtedly and necessarily neglect some components which each individual includes in that complex response to which he attaches the words socio-economic status. The scale makes us aware of this fact. If we feel that our personal definition, including our feeling tones associated therewith, is the only proper or true definition of the phenomenon under consideration, then the scale is, of course, invalidated by any disagreement on its part, at least to the degree that it differs from our common-sense discrimination. But this agreement of a measuring device with either an in-

---

which are so highly correlated with the way of doing the particular acts of the test, that the former can be predicted from the latter and vice versa. Under these circumstances, it is still true that the particular intelligence test tests all the correlated behavior as well as the particular acts involved in the test, hence the statement that intelligence is what the tests test. "What" here refers to *all* the certain ways of behaving which the test reliably discriminates and names in terms of scores. Intelligence, then, is the symbol which we employ to designate a certain aspect or evaluation of behavior, regardless of the particular acts in connection with which it occurs. Tests which reliably discriminate degrees of this aspect of behavior are called intelligence tests and what they test is called intelligence. For a fuller statement of this position, see my *Foundations of Sociology*, (Macmillan, 1939), pp. 446-57.

dividual or a consensus of common-sense judgments is not the only possible or even a primary object in scale construction. Scales may aim instead merely at *the reliable discrimination of any conditions whatever* which enable us to make successful adjustments. Sometimes these conditions as discriminated by instruments are at wide variance with the findings of common-sense. In many such cases we find it extremely wise to ignore the findings of common-sense and to follow those of the instrument. This is likely to be increasingly true as we have to adjust to remote and complex environments.

Our preference for the "qualitative" or prescalar use of such a phrase as socio-economic status is undoubtedly attributable to the fact that no one is compelled to break it up into components or to consider the relative weights he gives to each factor. The whole procedure is a deliciously private and subjective reaction for which we are not consciously accountable to anyone. Consequently, we feel strongly that the meaning we attach to such words has some intrinsic validity or fitness, as, for example, when children (and others) are impressed with the peculiar fitness of the word "cold" to designate a certain degree of temperature. For people reared in the same culture, and who receive their language-values from a fairly uniform source, there will be a certain uniformity in the use of such terms, usually sufficient to serve most primary-group purposes. When we come in contact with people of different culture backgrounds, and therefore with different word meanings, we marvel at their gross misjudgment of such matters as socio-economic status, living wage, decency, and so forth, and bring to bear against them unflattering emotions.

As science advances, we find less and less interest in such questions, for example, as "what" electricity *is*. Most people find it sufficient to define what electricity *is* in terms of what it *does*. It is *that which* under certain circumstances kills people, makes trains go, flashes in the clouds, illuminates lamps, makes the hand of the voltmeter move to a certain point, etc. As social science advances, we shall doubtless also find this *type* of answer adequate for the question as to what socio-economic status *is*. We shall be content to say that it is

that which under certain circumstances makes people beg on streets, fawn before the local banker, and behave arrogantly to a janitor or waiter. We shall say it is *that* status which is associated with certain kinds of houses, food, clothing, education, occupation; more specifically, we shall probably say that a person will be accorded status, *i.e.*, people will behave toward him *according to their estimate of the probability that he will achieve* the maximum goals of socio-economic striving. That is what we have meant by the term in pre-quantitative days; it is likely to continue to be what we mean by it under a quantitative terminology, except that we shall state it in terms of a number of units on a scale. That is in either case the only possible answer to the question of *what* socio-economic status *is* or *what* a socio-economic scale *measures*.

In short, it is only when we have a quantitative scale for measuring socio-economic status that we can give an explicit account of what we are measuring. We can enumerate or point to the items which enter into the construction of the scale and the proportional weight which each item is allowed in the total score. This is not possible when the general dictionary definition or folk usage is allowed to prevail and determine its meaning. We can never know whether another person means quite the same by the phrase as we do. When we try to determine this, we find ourselves constructing, formally or informally, a quantitative scale.

J. CONCLUSION

In this and the preceding chapter we have been concerned with some of the more recent and refined instruments for the measurement of social phenomena. The examples offered are, of course, merely illustrations of the large number of indices, scales, and other devices that have already been invented for the more objective study of social behavior. The examples we have used in the present chapter have been concerned more with non-verbal data, as compared with the attitude tests considered in the preceding chapter. These two types of instruments are not fundamentally different, however, as attitudinal behavior is always an aspect of what we have called institutional behavior. Conversely,

the "material" culture which has figured heavily in the scales considered in the present chapter, has been used as an index of attitudinal behavior. Socio-economic status, for example, might be studied either by attitude scales of the type discussed in the preceding chapter or by instruments like the Chapin and the Sewell scales. The accompanying bibliographical references give some idea of the number and variety of instruments that have already been invented for this type of social measurement.

We have seen that the same needs which have given rise to the various measuring instruments of physical science, the system of grading goods, services, and achievement, are gradually forcing the development of scales for the measurement of social institutions. The history of these sociometric scales is the same as that of all measuring devices—a gradual transition from purely subjective and qualitative judgments to increasingly objective and quantitative methods.

One of the principal obstacles to the development of such scales in the social sciences has been the widespread feeling that the services and values of many social institutions are so "intangible" or "spiritual" that they cannot be measured. We have shown, however, that all these values, if they are socially significant, have objective manifestations subject to quantitative measurement and comparison. It is true that for many forms of social behavior we do not as yet have standardized measures. This is the second great obstacle to the construction of valid sociometric scales. But their development must come through the constant revision and improvement of tentative measures constructed from the combined judgment of persons best qualified to judge. We have discussed methods of selecting such persons, and the principles which should govern the selection of units, standards, and weights. Finally, we have considered some concrete examples of the application of these principles, and the resulting sociometric scales.

These attempts represent a recognition of the crucial role of measurement in all science, as argued in the opening chapters. The undeveloped state of technics and instruments for this purpose in the social sciences is unquestionably a major obstacle to efficient and economical research in

these fields. The matter can perhaps be summarized most briefly by imagining what would be the state of knowledge, research, and practical efficiency in the physical sciences if every researcher had to invent anew for each research project, instruments for the measurement of time, length, weight, temperature, etc. *The corresponding state of affairs is virtually what today exists in the social sciences.* There is no systematic theory which stipulates what we choose to regard as fundamental dimensions of social behavior and there are few standardized and verifiable definitions and measures of the various concepts and dimensions used by various researchers.[45] Deliberate and coordinated work on this key problem might properly engage some of the attention of research foundations interested in fundamental research.

K. SUGGESTIONS FOR FURTHER STUDY

1. Perhaps the best summary and bibliography in existence of measuring instruments of the type considered in this chapter is Wm. H. Sewell, *The Construction and Standardization of a Scale for the Measurement of the Socio-Economic Status of Oklahoma Farm Families,* Oklahoma Agricultural and Mechanical College, Agricultural Experiment Station, Technical Bull. No. 9, April, 1940 (Stillwater, Oklahoma). Part I of this excellent bulletin summarizes the story of some twenty-five attempts during the decades of 1910–40 to measure socio-economic status. The rest of the bulletin gives in detail the steps in the construction of such a scale. The entire bulletin is recommended as supplementary reading to the present chapter.

2. The details in the construction of S. C. Dodd's scale for measuring rural hygiene in Syria have been mentioned in several previous connections. Part I of his book, *A Controlled Experiment on Rural Hygiene in Syria,* (Oxford Univ. Press,

[45] This work will necessarily, to begin with at least, be heuristic in nature, *i.e.*, different students will propose different theories, concepts, and measures, of which the most useful will survive in part or whole. An example of one of the most comprehensive and systematic of such attacks is found in S. C. Dodd, *Dimensions of Society,* (Macmillan, 1941).

1934), remains a classic example of the subject discussed in the present chapter.

3. For a detailed discussion of the methods and problems of constructing a sociometric scale with special reference to the measurement of morale, see E. A. Rundquist and R. F. Sletto, *Personality in the Depression,* (Univ. of Minn. Press, 1936).

4. An excellent discussion of the tests of significance will be found in J. H. Smith, *Tests of Significance: What They Mean and How to Use Them,* (Univ. of Chicago Press, 1939).

5. Three important indices for measuring rural planes of living from census and other published sources of information are the following:

a. C. E. Lively and R. N. Almack, *A Method of Determining Rural Social Sub-Areas with Application to Ohio,* (Department of Rural Economics, Mimeograph Bulletin 106, Ohio State Univ., Columbus, O., 1938).

b. M. M. Blair, *Indices of Levels of Living for the Thirteen Southern States by Counties,* 1930, (Oklahoma Agricultural and Mechanical College, Social Science Research Council, Bull. No. 2, Stillwater, Okla., 1939).

c. C. Goodrich, B. W. Allin, and M. Hayes, *Migration and Planes of Living,* (Univ. of Pennsylvania Press, Philadelphia, 1935).

6. A very useful book of suggestions as to items of importance in different aspects of community life is Joanna C. Colcord, *Your Community,* (Russell Sage Foundation, 2nd Edition, 1939).

7. Some ingenious suggestions for the development of scales for measuring the degree to which motion-picture plays violate community mores may be found in C. C. Peters, *Motion Pictures and Standards of Morality,* (Macmillan, 1933).

8. For a type of social measurement not treated in this book, but which is of basic importance, see G. K. Zipf, *National Unity and Disunity,* (Principia Press, Bloomington, Ind., 1941). See also, my *Foundations of Sociology,* (Macmillan, 1939), Chaps. 9-13.

CHAPTER X

SOCIOMETRIC METHODS IN ECOLOGY AND INTERPERSONAL RELATIONS

Perhaps the relations of a society, a social system, and the individuals who compose the society can be made clearer by an analogy. A rope can be analyzed into its component strands with respect both to its total length and to its content at any point in that length. Its structural pattern can also be ascertained by observing the spatial relations of the various strands at a series of points along its length. This pattern will persist in spite of the termination of certain strands and the introduction of others and will bear little relation to the individual qualities of the various strands — such things as their exact length, thickness, or color. It can be described in the abstract and compared with the structural patterns revealed by other ropes. A society, in its extension through time, can be likened to a rope braided from the short strands of individual lives. By studying the social relations between individuals and observing the repetitive situations, it becomes possible to deduce the structural pattern of the society. This pattern tends to persist in spite of the steady turnover in the society's content and bears little relation to the special qualities of the various individuals who occupy places in the society at various times. It can be described in the abstract and compared with the structural patterns of other societies. — *Ralph Linton.*[1]

A. TYPES OF SYMBOLIC REPRESENTATION OF SOCIAL PHENOMENA

The social sciences are concerned with the explanation of social groupings and the interactions that are called group behavior. To carry on this work it is necessary, as we have seen in previous chapters, to develop objective methods of depicting symbolically the phenomena in which we are interested. In order to emphasize that all of the detailed methods we have considered above are merely special techniques in this comprehensive enterprise, and in order to illustrate the relationship of all of these methods, we shall summarize in the present chapter some of the principal types of objective symbolic representation. We find it convenient to discuss and illustrate this development in four general categories.

1. Geographers and ecologists were among the first to develop objective symbolic methods regarding certain aspects of human groupings. Geographic and ecological maps show human relationships with special reference to their distribu-

[1] "A Neglected Aspect of Social Organization," *Amer. Jour. Sociol.*, 45:870-86, 1940, p. 871.

tion and function in geographic space. In its more elementary forms, this type of symbolic representation is highly objective and generally understood because it is almost a kind of picture-writing. Also, cartographic methods of representation have hitherto been applied chiefly to those aspects of the community which are conspicuously incorporated in the so-called material culture, for the objective study of which our symbolic techniques have always been more highly developed. The representation on paper of such data as area, direction, topography, channels of communication, land utilization, economic areas, etc., is, therefore, relatively standardized and generally understood. As a result, we ascribe to those aspects of the community which can be depicted or represented by the simpler cartographic methods a "tangibility," "concreteness," and "reality," which we withhold from aspects for which we have not yet developed objective symbolic techniques.[2]

2. After mere mapping, the next step in the process of describing the community was to convert behavior itself into symbolic indices which can in turn be represented in geometric, isometric, isotypical, or gradient terms.[3] Thus birth, death, disease, and delinquency rates, or rates for any other behavior phenomena can be computed for local areas and the relationship of such phenomena to these regions can be shown by gradient lines or other geometric devices. A line on the same map representing the declining delinquency rate from the center of the city toward the periphery, denotes a phenomenon just as real, observable and meaningful as a line representing the main street. (See Figures 10 and 12 below.) Yet such is the degree of our unfamiliarity with such abstractions as rates, that we feel the line representing the street has a certain "reality," legitimacy, and relevance to the "true" representation of a community which it is felt

[2] For a development of this position, namely, that tangibility, concreteness, etc., are not characteristics inherent in data but merely words describing the degree to which we have developed objective symbolic techniques of dealing with such data, see G. A. Lundberg, "The Thoughtways of Contemporary Sociology," *Amer. Sociol. Rev.*, 1:703-23, 1936. Also A. F. Bentley, *Behavior Knowledge Fact*, (Principia Press, Bloomington, Ind., 1935), Chaps. 19, 21.

[3] See C. R. Shaw, *Delinquency Areas*, (Univ. of Chicago Press, 1929). Elsa S. Longmoor and E. F. Young, "Ecological Interrelationships of Juvenile Delinquency and Population Mobility," *Amer. Jour. Sociol.*, 1:598-610, 1936.

the delinquency gradient has not. Both types of lines represent both a structural and a functional phenomenon. At present many of these behavior phenomena have not yet been metricized. Until they are reduced to metric units we may deal with them by topological constructs, or in terms of symbolic logic. This may be an intermediate step toward metrical statement.[4] The correlation[5] of these various indices, when worked out for *all* the relevant aspects of community behavior, would represent a comprehensive description and explanation of the life of a community.

3. The aspect of community life that this approach seems to neglect is the so-called "human" psychological factor of social interaction itself. Hence the next step is to invent objective devices for the representation of these processes which are at present dealt with largely in the language of literature and philosophy.

It will be observed that the transition in the technique of representation from an ecological spot or rate map of delinquency to the organization or functional chart represents merely a transition (*a*) from geographic to social space representation, and (*b*) from the relatively pictorial technique of the delinquency map to a more abstract form of symbolic representation of the behavior depicted by the functional chart. That is, the lines in an organization chart do not correspond to roads, railways, etc., in geographic space, but represent relationships in social space,[6] such as the relation of the mayor to subordinate officials, just as a political boundary or a delinquency gradient represents a behavior phenomenon relating two areas of a city in geographic space. In both cases the lines stand fundamentally for behavior phenomena of concern to social science. In the same way, purely topological representations or the characters of symbolic

[4] See J. F. Brown, *op. cit.*, 476 ff. Also K. Lewin, *Principles of Topological Psychology*, (McGraw-Hill, 1936), Part II; and especially Dodd's *Dimensions of Society*, (Macmillan, 1941).

[5] The term correlation is here used in its broadest sense. See G. A. Lundberg, "Quantitative Methods in Social Psychology," *Amer. Sociol. Rev.*, 1: 41-43, 1936.

[6] *Cf.* J. F. Brown, "On the Use of Mathematics in Psychological Theory," *Psychometrika*, 1:77-90, 1936. See also, P. A. Sorokin, *Social Mobility*, (Harper, 1927), Chap. 1, for a good non-mathematical statement of the same point. Algebraic description should, of course, supersede the geometric as soon as possible.

logic are merely other literary methods of representing relations. These points are made for the benefit of those who feel that any other language than the customary philosophical terminology of discussing the subtler community relations is "artificial," "cold," rigid, and lacking in color. All symbolic representation of behavior is "artificial" in exactly the same way. The coldness, rigidity, and lack of color of the topological and mathematical symbolism may be the chief reason for its superiority as scientific language. It is, in fact, because philosophical vocabulary is so suffused with subtle and varied emotional tones and meanings that it turns out to be so often inferior in scientific discourse and hence is supplanted by geometric and other mathematical symbols. It is significant to note how well other sciences get along in explaining the most intricate phenomena without any of the type of language alleged to be so essential in sociology.

Very similar to the formal, official, or latent organization of the whole community is a network of so-called private groups—families, religious bodies, clubs, fraternities, etc. These have a considerable degree of formal organization and permanence, but are usually thought of as relatively selective and "voluntary" instead of inclusive and automatic in their membership.[7] Also they are more limited and specialized in the type of interests which give rise to them. This level of community organization is usually regarded as less "tangible" than the formal political organization of the whole community. The feeling that these informal groupings are less tangible is due to the relative absence of objective and systematic information about them and the relatively unstable and shifting nature of most of them. Attempts to study and chart this level of community structure are, however, becoming increasingly objective.[8] This is notably true

[7] The family is probably an exception to this statement. But it is for present purposes classified among the so-called voluntary associations arising from drives of relatively widespread and permanent distribution, and hence sociologically comparable to the religious, fraternal, recreational and socio-economic organizations that make up the bulk of the "voluntary" associations of a community.

[8] See Mirra Komarovsky, "A Comparative Study of the Voluntary Organizations of Two Suburban Communities," *Publ. Amer. Sociol. Soc.*, 27:83-93, 1933. Also, by the same author, *The Unemployed Man and His Family*, (Institute of Social Research, New York, 1940). Also, G. A. Lundberg, *et al.*, *Leisure*, (Columbia Univ. Press, 1934), Chap. 5; F. S. Chapin, *Contemporary American Institutions*, (Harper, 1935), Chaps. 16, 18.

of studies of family structure, function, and behavior.

4. Finally, there is, underlying these private and more or less "voluntary" social groupings, the tremendously intricate and elementary network of informal social nuclei which we describe in such terms as affinities, friendships, love affairs, cliques, gangs, etc. These are at present the least tangible social units because of their informal, transitory, private and frequently secret character. They may even be illicit or illegal in nature. These groupings are not structurally embodied in any of the objective forms usually accompanying other forms of social organization, such as written constitutions, buildings, and other "material" paraphernalia, and they are charged with no formal functions by the larger community. Yet the basic nature of these nuclei and their profound influence in determining especially the functioning of the more formal structures has always been recognized. The influence of these alignments is what we rely upon "insiders," "dopesters," and gossips to convey to us, since objective and official accounts of the behavior of the generally recognized social groups contain no mention of these underlying influences. Yet it is admitted that the generally recognized behavior of a community can in a given case be adequately explained only if these basic alignments are understood. Behind all present objective explanations of community behavior, therefore, there is always an aura of data, at present largely in the form of gossip, because of its subjective form, which we all feel constitutes the "real" explanation of what takes place. In the same way, the clever leader knows that the success of any program must take into consideration the cultural, psychological, emotional, and other clique alignments, especially of "key" people, the existence of which may not be consciously recognized by the people concerned, and certainly not admitted by them. Failure to take these alignments into consideration is generally recognized as a tactical blunder, lack of tact, diplomacy, etc. The objective study of these elementary social nuclei[9] is, therefore, quite basic

[9] The concept of social nucleation has hitherto been employed in ecology chiefly to refer to that process through which centers of interaction develop in social aggregates, such as metropolitan areas and geographic regions. Within a region or a metropolitan area, for example, "the population tends to subdivide and become multinucleated in a complex of centers that are

to an understanding and intelligent control of the community.

For purposes of illustration we shall group the above classifications into two main categories: (1) those having to do with depicting social phenomena in geographic space, or the techniques of human ecology; and (2) those having to do with the fundamental structure of human groups from the standpoint of interaction.

B. TECHNIQUES OF HUMAN ECOLOGY

One of the most elementary and common methods of representing the relationship of social phenomena in geographic space is the ordinary political map. It shows the concentrations of people in cities, administrative areas, and the chief avenues of communication. When we are interested in a more accurate representation of the frequency of recurrence of certain phenomena, the familiar spot-map or cross-hatched map is used. The latter type of map shows the degree to which a stated phenomenon characterizes a given area, and such maps are the commonest and perhaps the clearest way of depicting cultural regions.[10]

Since cross-hatched maps usually represent *rates* rather than raw data, these maps can be used only when the data necessary for the computation of rates are available. A spot-map of juvenile delinquency, for example, may be made simply by placing a dot at the address of the delinquent. But if it is desired to compare different areas as to their delinquency-producing character, we obviously cannot compare merely the total number of delinquents in each area, if one area has, for example, ten times as many juveniles as the other. The *rate* of delinquency, not the absolute numbers, is in this case the significant comparison. That is, we are interested in the relative number of delinquents in the two areas, *per hundred, per thousand,* or some other number-unit of juveniles in the population.

---

economically and socially integrated in a larger regional unity." R. D. McKenzie, *The Metropolitan Community,* (McGraw-Hill, 1933), p. 49. F. S. Chapin also uses the term to describe certain institutions as "nucleated" (family, church, etc.) as contrasted with "diffused-symbolic" institutions (art, law, science, etc.). *Op. cit.,* Chap. 2.

[10] Excellent illustrations will be found in Howard W. Odum and H. E. Moore, *American Regionalism,* (Holt, 1938).

Rates are one of the commonest of statistical devices and in their simpler and commonly experienced forms are quite generally understood. But the selection of a satisfactory base for the computation of rates is frequently a very difficult matter. The basic formula for a simple rate is

$$\text{Rate} = \frac{\text{The number of times a specific kind of event occurs in a given period}}{\text{The whole number of exposures to the risk of its occurrence}}$$

To choose a proper denominator in this formula we need to know with what component of a situation the phenomenon we are studying is most stably and significantly correlated. This is why "the number of married women aged 15 to 45" is a better base for the computation of a birth rate than "total population" or some largely irrelevant or illogical base such as "national income," "total bushels of wheat produced" or some equally unrelated phenomenon. There are many social events the frequency of which can be counted with the same accuracy as births. But for many of these events we do not know what common component of the situation is so highly and meaningfully associated as is *births* with the *female population aged 15 to 45*. Dependable rates, therefore, can be computed only after a good deal is known about the *relationship* of the phenomenon to other components of the situation.

The simplest form of rates is found in such expressions as "miles per hour," "persons per [one] square mile," "income per [one] year," etc. In such cases as "births [per year] per thousand population," "income [per year] per capita," the numerator is itself a rate although in these cases we are so accustomed to counting such events as births and income payments on a yearly basis that the time unit is usually not mentioned. The computation and interpretation of rates is a subject of great importance which we must leave to the proper statistical treatises.[11] We have referred to them here only because they are basic to the representation of social phenomena in their geographic relationships as depicted by ecological maps, which are here our chief interest. The com-

[11] For a good brief treatment, see F. E. Croxton and D. J. Cowden, *Applied General Statistics,* (Prentice-Hall, 1940), Chap. 7.

putation of a significant rate may involve all the techniques described throughout this book, plus statistical operations which we have not considered. Ecological maps are devices for the display of *the relationships between areas with regard to the characteristic expressed by the rate.* In short, ecological maps represent a method of correlating certain phenomena on a geographic basis.

What has been said above regarding the use of rates in ecological maps also applies to the use of indices of social conditions as, for example, the indices of socio-economic status, efficiency of health departments, sanitary conditions, and others discussed in the preceding chapter. An index is a more general designation than a rate, consisting frequently of a combination of rates or other figures representing degrees of phenomena in which we are interested. Thus, the Work Projects Administration in its division of the United States into rural cultural regions and sub-regions (see Figure 9) computed the following indices for every county in the United States: [12]

(1) A rural-farm plane-of-living index combining the average value of the farm dwelling, the percent of farms having automobiles, the percent of farm homes having electric lights, the percent having running water piped into the house, the percent having telephones, and the percent having radios, 1930.[13]
(2) A rural-farm population fertility index constructed by computing the ratio of children under 5 years of age to women 20-44 years of age, 1930.
(3) Percent of farms producing less than $1000 gross income, 1929.
(4) Percent of farm tenancy, 1935.
(5) Land value per capita of the rural-farm population, 1930.
(6) Percent of farm produce consumed on farms, 1929.
(7) Percent of rural families residing on farms, 1930.

The factors of more localized importance were:

(1) Percent Negroes constituted of the total rural-farm population in the South, 1930.
(2) Percent "other races" constituted of the total rural-farm population in the Southwest, 1930.

[12] A. R. Mangus, *Rural Regions of the United States,* (Works Projects Administration, Division of Research, 1940), pp. 79-80.

[13] C. E. Lively and R. B. Almack, *A Method of Determining Rural Social Sub-Areas with Application to Ohio,* (Mimeograph Bull. No. 106, Ohio State University and Ohio Agricultural Experiment Station, Columbus, Ohio, January, 1938).

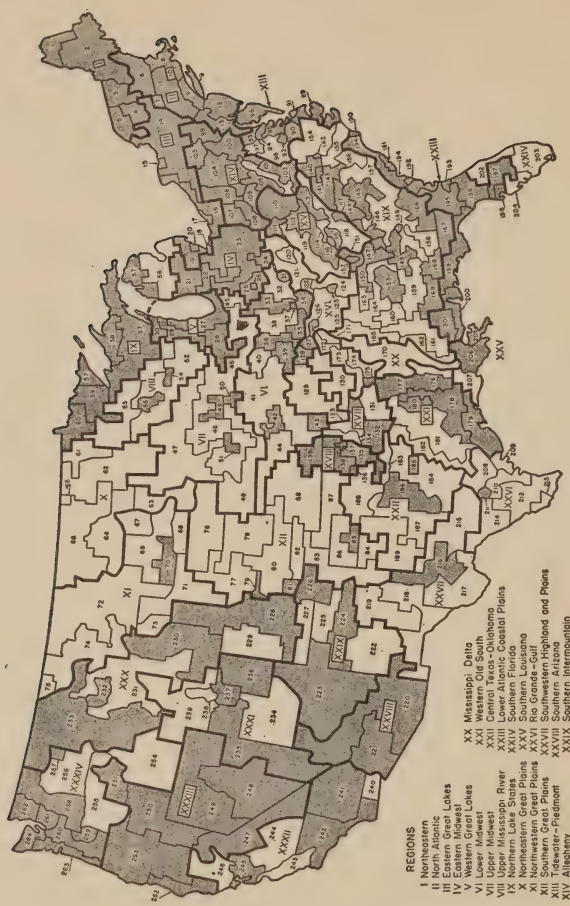


FIGURE 9. Rural Cultural Regions.

From Rural Regions in the United States. By A. R. Mangus (Works Projects Administration 1941).

In addition, physiographic features were taken into account in places where these were prominent elements.[14]

Some people may object to this procedure on at least two grounds: (1) Not all of the factors are cultural in the narrower meaning of that term, and (2) so many-sided a phenomenon as the culture of an area cannot be adequately characterized by ten indices alone. Both objections are fundamentally of the same character, and spring from a fail-

[14] A number of additional factors were used to describe the individual regions. See A. R. Mangus, *op. cit.*, Tables 2 and 3.

ure to recognize that the innumerable individual traits of the physical as well as the social world are nevertheless *highly related to each other in enormous clusters or constellations of traits.* To the extent that the intercorrelation of traits in each constellation is high, a single one of these traits may be used as *an index* of the whole cluster. It is this fact which makes such a device as an index possible, as we showed in the preceding chapter. It is by virtue of this fact, also, that we may frequently infer and measure cultural situations in terms or by means of non-cultural indicators. If the ten indices enumerated above are *sufficiently highly and reliably correlated* with most or all of the other phenomena which we have in mind when we speak vaguely of the "many-sidedness," the "unique quality" and "spirit" of "a culture," then these mundane indices are, for scientific purposes at least, perfectly valid devices for designating and defining culture areas. When these areas are depicted on a map, their geographic relationship to each other and to other features indicated on the map becomes apparent. The discovery of such relationships is the principal quest of science in all fields.

Numerous studies have been made to determine to what extent such indices as those listed above are correlated with each other in different parts of the United States. One study[15] took a preliminary list of 83 factors and through an extensive but incomplete factor analysis found that at least one-fourth of the 83 factors were correlated to the extent of .60 or above, with three variables, namely, a plane of living index, a population fertility index, and a farm income index. Among the factors associated in a high degree with these three indices were such socially significant variables as migration, age distribution, infant mortality, size of farm, farm tenancy, mortgage debt, part-time farming, one-room school enrollment, State school aid per pupil, and churches with full-time pastors.[16]

Commenting on the study resulting in the identification of the regions shown in Fig. 9, A. R. Mangus says:

> We are told that the regional distribution of primitive American culture was discovered when archaeological specimens in the

[15] C. E. Lively and R. B. Almack, *op. cit.*
[16] A. R. Mangus, *op. cit.*, p. 79.

American Museum of Natural History were grouped according to their similarity and spotted on a map by place of origin.[17] Similarly, it was found that when cards containing county ratios representing broadly significant factors are brought together on the basis of homogeneity, they represent counties that tend to fall together or in the same general locality. Moreover, they often represent counties that are known to be located within a distinct culture area. . . Those counties with a low plane-of-living index and simultaneously low population fertility, delimit the Mississippi Delta with its one cash crop (cotton) and its high incidence of Negroes and tenants. The extreme opposite of the Delta country is found in Utah and Idaho, where a high plane of living is associated with a high ratio of children to women, and where a distinct culture based on religion is present. On the basis of analyses of this kind the conclusion was reached that the selected factors have sufficiently wide significance to serve as determinants of what may properly be called sociological or cultural regions and subregions.

The division of cities into "natural areas" has also been carried out according to the same basic method described above. The tabulation of census figures for small local areas (census tracts),[18] which has been greatly extended in the last decade, enormously facilitates the computation of more sensitive indices than was true when such figures were available only for wards or other relatively large and non-homogeneous areas. The major types of areas in a city, such as "gold coast," apartment house areas, rooming house areas, slums, business sections, warehouse areas, etc., can roughly be located, of course, without detailed statistical data, although these areas frequently shade into each other and can be characterized as of one or the other type only in degree. For less conspicuous characteristics of local areas, however, we are dependent on rates and indices, for example, of mortality, morbidity, delinquency, crime, nativity, education, and all the factors entering into socio-economic status. The possibility of computing rates for such phenomena for small areas, sometimes including only a few city blocks, has made possible a number of significant researches into urban patterns. One of these patterns is illustrated in Figure 10.

[17] Alexander A. Goldenweiser, "Cultural Anthropology," in *History and Prospects of the Social Sciences,* edited by Harry Elmer Barnes, (New York : Alfred A. Knopf, Inc., 1935), p. 244.

[18] The Census of 1940 will be tabulated according to census tracts for sixty cities.

FIGURE 10. Rates of Delinquency Cases by Mile Zones Surrounding the Loop (Chicago).

In this illustration, cases were first plotted on a base map according to home addresses, and rates were then computed for each census tract (not shown in Figure 10) using the total male population aged 10 to 16 years in each tract as a base. Rates for the square-mile areas were computed on the basis of population estimated from census enumeration districts and census tracts. Rates for each zone were secured by averaging the rates for each square-mile area falling within each zone.[19] The consistently declining rate of juvenile delinquency as we proceed outward from the "Loop," illustrates the *gradient* in social phenomena. A gradient represents a gradation in intensity—rate of change—in a social condition, varying with its distribution over a given area. Such gradients can readily be illustrated by drawing radial lines from the central business district of a city in different directions and computing the rate (birth, death, delinquency, socio-economic status or any other characteristic) for each census tract touched by the radial line. To the extent, as in the case illustrated, that rates vary consistently in any observable way, at least an hypothesis emerges as to possible laws or generalized patterns of urban development.

If similar studies in other cities confirm the findings, we may generalize and predict from such patterns. The usual reservation of "other things being equal," necessary in all scientific generalizations, applies, of course, also to generalizations regarding urban patterns. It is not necessary to assume that the pattern illustrated above is everywhere present. The technique described is equally useful for detecting and explaining *whatever* patterns are found, or in measuring deviations from any real or theoretical pattern we may adopt as a type. Researches on urban communities employing the methods described above are admirable illustrations of a type of scientific analysis. If, in addition to the conventional demographic and social work data which are at present available, similar analyses were undertaken regarding the distribution of opinions, attitudes, and the subtler social characteristics of individuals and local communities,

[19] C. R. Shaw, *Delinquency Areas,* (University of Chicago Press, 1929), p. 25 ff. On the statistical significance of small but consistent differences between zones, see C. C. Peters, "Note on a Misconception of Statistical Significance," *Amer. Jour. Sociol.,* 39:231-36, 1933.

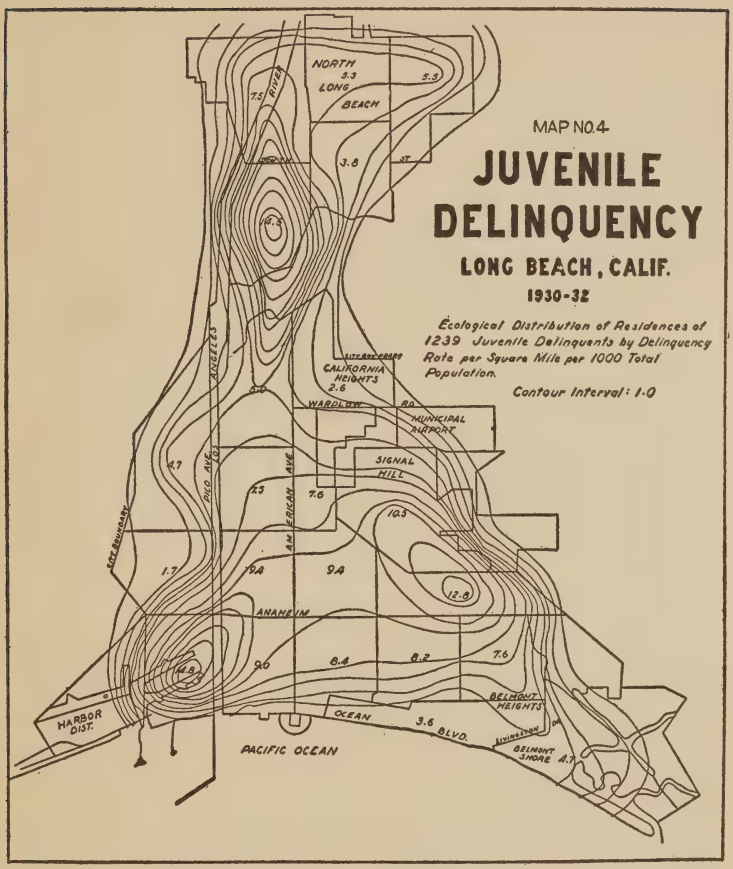


FIGURE 11. Juvenile Delinquency Map.

illuminating generalizations on urban communities might well result.

A variation of the method of radials and zones is the *isometric map.* Areas of equal rates are connected by lines in the manner that geographers represent varying altitudes and meterologists indicate areas of varying atmospheric pressure. (See Figure 11.) This technique avoids the impression of abrupt changes in rates at the boundaries of areas, which is

suggested by the ordinary cross-hatch map. It is true that the simple spot map also shows correctly the gradual change in the prevalence of some phenomena, but gives no indication of the population base. The isometric map is designed to correct both of these shortcomings. In the isometric map convenient points in the range of rates from the highest to the lowest are located by interpolation and connected by lines as shown in Figure 11.[20] Areas bounded by any two isometers then represent belts within which the phenomenon charted shows a constant rate of occurrence.

There are many ecological techniques which have not been mentioned above.[21] All of them depend ultimately upon the location of phenomena in geographic space. Any organization or instruments which facilitate this task will make possible scientific advance on a very important front.

The most important single development in this direction is perhaps the widespread adoption of *census tracts* as a basis for gathering and registering social statistics. The spotting of cases in their proper geographic location is greatly facilitated by a street index or coding guide which includes all the addresses in the city classified alphabetically by streets, and showing what house numbers on each street are in a given census tract. Schmid[22] suggests that the greatest speed and accuracy in locating cases according to addresses can be obtained by writing all the addresses to be spotted on separate cards and sorting them in alphabetical order (or serially for numbered streets) and then arranging the cards in the order

[20] For a detailed account of methods of constructing such maps the student should consult E. R. Mowrer, "The Isometric Map as a Technique of Social Research," *Amer. Jour. Sociol.*, 44:86-96, 1938. See also, Elsa S. Longmoor and Erle F. Young, "Ecological Interrelationships of Juvenile Delinquency, Dependency, and Population Mobility," *Amer. Jour. Sociol.*, 41:598-610, 1936.

[21] For the best brief summary of these methods, see Calvin F. Schmid, "The Ecological Method," (Chap. 14 in Pauline V. Young, *Scientific Social Surveys and Social Research*, Prentice-Hall, 1939). Special attention should be called to Schmid's discussion of centrographical and related methods of determining the center of mean distances of spatial distributions, (*e.g.*, the changing "center of gravity" of population distribution or other data). For more extended treatment of this subject, see "The Centrographical Method and Regional Analysis," by E. E. Sviatlovsky and W. C. Eells, *Geog. Rev.*, 27:240-54, 1937. See also, Louis Wirth and Margaret Furez (Editors), *Local Community Fact Book*, 1938, (Chicago Recreation Commission, 1938).

[22] *Op. cit.*, p. 391. For a good illustration of one such index, see Calvin F. Schmid, *Census Tract Street Index for Minneapolis, Minnesota*, (Minneapolis Council of Social Agencies, 1936).

of the house numbers on each street. Large numbers of individual addresses can thus be located in their proper census tract by a single reference to the index. In the meantime, increasing amounts and kinds of social statistics will doubtless be recorded and tabulated by census tracts, thus permitting the computation of rates by these small districts for all data so recorded. There is no doubt that the method affords tremendous possibilities for significant sociological analysis and for rigorous verification of sociological hypotheses which has hitherto been lacking in much social research.[23]

C. SOCIOGRAPHIC METHODS

It has been noted that most of the ecological methods here described depend upon the availability of demographic statistics for given geographic areas. The accumulation of these statistics may involve any or all of the methods described in this book. The techniques of discovering and exhibiting the relationships of these data in geographic space are the additional features introduced by ecological methods. Broadly speaking, these methods are graphic devices for correlating social phenomena with the factor of geographic space. The charting of configurations of phenomena in time and space is perhaps the most basic general task of all science. The organism's orientation to its geographical-spatial environment is undoubtedly one of the earliest which it acquires, and is one of the most elementary of its adjustment techniques. It is not surprising, therefore, that relational thinking in every field and on every subject tends to structure itself in spatial terms.

Thus it happens that the description of purely social relationships in spatial terms is common in perhaps all folk languages. That is, it is very common to speak of high and low status in life, one person being below another, of social mobility, of social distance, social boundaries, social barriers, etc. These terms in such contexts clearly have no reference at all to geographic space. Yet they are just as meaningful ways of communicating social relationships as are lines on a

[23] For a striking use of census tract data in this way, see S. A. Stouffer, "Intervening Opportunities: A Theory Relating Mobility and Distance," *Amer. Sociol. Rev.*, 5:845-67, 1940.

topographic map in communicating relationships of geographic space. Hence it is, also, that many a lecturer on the most abstract philosophical subjects will resort to spontaneous drawings on the blackboard in order to communicate more adequately certain relationships which have no geographic *loci* whatever. Certain branches of mathematics consist, indeed, of entire systems of such space-representations in which positional relationships of any kind may be expressed. We have considered above, methods of representing the relationships of social phenomena in geographic space, after the manner suggested by plant and animal ecologists. In turning to the methods of representing social relationships in spatial symbols and terms, therefore, we are not imposing a geographic *analogy* upon the social world. We merely adopt spatial symbols to represent *relationships* regardless of the types of subject matter among which the relationships exist.

The commonest cartographic device for representing social relationships is perhaps the familiar organization chart. A city government, for example, is represented by a series of circles or rectangles connected by lines. A rectangular box at the top represents the city manager or the mayor; another box on the same horizontal level represents the city council; below each are smaller rectangles representing officials "under" each of the above; lines from each figure on the first level show whether the mayor or the council controls each of the respective "subordinates," etc.

More complicated social relationships constituting other aspects of social structure may be represented as in Figure 12. This diagram shows how the status of men and women in one society varies with age, sex, and the various personal and occupational characteristics represented by each of the persons here charted. Different ways of achieving equal status is apparent. For example, a female child (3) aged 10 must be an "honored-child" to approximate the status of any boy (11). Likewise, an adult woman (7), if she is notably older, and a medicine woman, may surpass in status a mere adult male (13) unless he is a war leader (14). Also a medicine woman (7) approaches the status of a medicine man (15) only if she is notably older. If the horizontal axis of this dia-

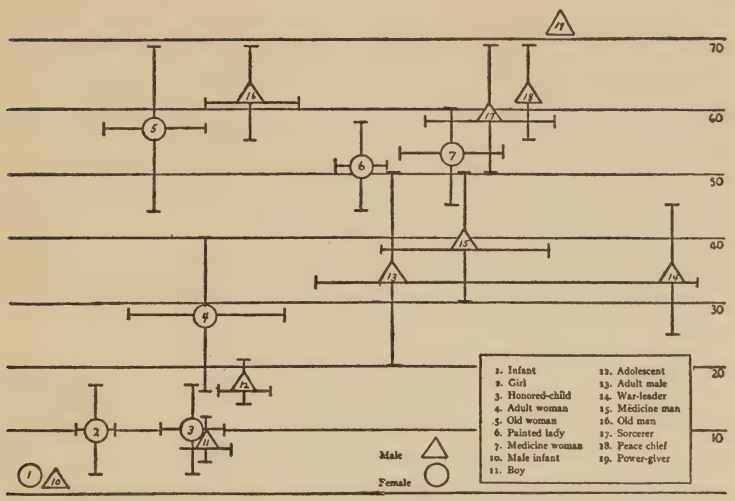


FIGURE 12. Comanche social structure (From Linton [24]) "Age levels are represented by the long crosslines. The vertical lines passing through either triangles or circles indicate, by their length, the total age range of individuals occupying the status. The position of the triangle or circle with relation to the total length of this line does *not* indicate the mode of age distribution among the individuals occupying the status, although this refinement could easily be introduced in cases where vital statistics were available for the society. The horizontal lines connected with circles or triangles indicate the range of prestige among individuals occupying the status. Similarly, the horizontal organization of the entire figure is based on the relative prestige of the various statuses represented."

gram consisted of a status scale in standard units, these relationships could, of course, be represented by crude equations resembling ordinary mathematical expressions. Similar figures could be constructed for any other coordinates as well as for age and status. Multi-dimensional figures might take into consideration additional determinants.[25]

The relationships depicted in charts of this sort do not have to be metricized, however, in order to be meaningful or useful. Non-metrical charts and schematic diagrams have always been valuable devices for depicting relationships of

[24] Ralph Linton, "A Neglected Aspect of Social Organization," *Amer. Jour. Sociol.*, 45:870-86, 1940.

[25] See, for an example, F. S. Chapin, *Contemporary American Institutions*, (Harper, 1935), p. 352.

various sorts. As methods of *measuring* the relationships involved develop, and as we desire to generalize more and more broadly, we shall doubtless be forced here, as in other fields, toward more refined mathematical symbolization and toward increasing use of tested methods of calculating and stating averages, deviations, and probabilities. In the meantime, the diagrammatic methods are valuable analytical and expository tools in their present form.

Figure 12 depicts *one kind* of social relationship or structure, namely, prestige or status. Cartographic methods have also been used to represent attraction-repulsion patterns in groups. Moreno[26] asked all the members of the State Training School for Girls at Hudson, New York, to indicate their choices, in order from one to five, of the members of the community with whom each would like to occupy the same house. The choices were confidential, and it was definitely understood that they would become the basis for a reassignment to different houses. This is a closed community with a population of between five and six hundred. The girls live in sixteen cottages, each with a housemother. Tabulation of the choices of desired housemates, including housemothers, revealed an intricate set of societal nuclei of five principal types. (1) The most elementary and definite nucleus was the mutual first choice of two or more individuals among themselves. This might be represented by figures of couples, triangles, squares, circles, or figures of more intricate dimensions. (2) Then there were the patterns most simply represented by chains of a non-mutual character, as, for example, when A chooses B, B chooses C, and C chooses A. Such chains might, of course, consist of any number of individuals. (3) A third type of configuration resulted from a clustering

[26] J. L. Moreno, *Who Shall Survive? A New Approach to the Problem of Human Interrelations,* (Nervous and Mental Disease Publishing Company, Washington, D. C., 1934), 23 ff., 373 ff. See also, D. Thomas, A. M. Loomis, and R. Arrington, *Observational Studies of Social Behavior,* (Institute of Human Relations, Yale University, 1933); Ruth Arrington, *Interrelations in and Behavior of Young Children,* (Child Development Monographs, No. 8. Bureau of Publications, Teachers College, Columbia University, 1932); W. I. Newstetter, M. J. Feldstein, and T. M. Newcomb, *Group Adjustment. A Study in Experimental Sociology,* (Western Reserve Univ., 1938); K. Lewin, "Field Theory and Experiment in Social Psychology: Concepts and Methods," *Amer. Jour. Sociol.,* 44:868-89, 1940.

of a large number of choices around a single individual, to some of which this "star" may respond by mutual choice, to others by indifference, and to others by repulsion. (This information regarding indifference and repulsion was secured through separate private interviews.) (4) A fourth type of nucleation, of great practical significance, was the phenomenon of the powerful as contrasted with the popular individual, and the phenomenon of indirect influence. Thus an individual may be the object of a large number of choices by people relatively isolated from the rest of the community. The center of such a cluster of choices may be very popular with her immediate group, but may not have very much influence in the community because of the limited contacts of her admirers. On the other hand, an individual chosen by only a few may wield a powerful indirect influence by virtue of the extensive connections of the few popular persons for whom she is a center of attraction. Moreno cites one case, herself chosen by only four individuals, all of whose attraction she reciprocates. But these four are in their turn chosen by nearly one hundred individuals. This seems to be the sociological position of the "power behind the throne" who wields the real power through the personalities of popular nominal leaders. As Moreno has pointed out, "the distribution of power in large groups depends upon the intricate distribution of emotional currents. An individual who is in control and can steer the course of one of these currents can wield an immense potential influence out of all proportion to his immediate following." Finally, (5) there is the isolated individual, chosen by nobody, although he may choose some other persons.[27] Such individuals are almost certain to be maladjusted and to cause maladjustment in any group with which they have to live. Sometimes, to be sure, such individuals, not chosen by anyone as housemates, may, however, be chosen as desirable work companions, so that isolation is always with reference to a criterion. The completely isolated individual would be one who was not chosen by anyone as an associate in any of the activities or

[27] The five general types of grouping summarized above are, of course, subject to extensive variation, elaboration, and sub-classification. See Moreno, *op. cit.*, 114-33; 147 ff.

relations of a community. Such an individual could not be very sensitive to the behavior standards of the community because he is cut off from the currents which constitute the pressures by which these standards become operative on the individual, and hence he would almost certainly be a social problem.

Now the individuals attracted to another individual regarded as a nucleus,[28] together with the individuals to which this nucleus is attracted, any or all of which attractions may or may not be mutual, may be regarded as a social atom[29] with reference to any social interest, *i.e.*, any adjustment need. Thus, each individual, provided he is not socially isolated, is the nucleus of one or more social atoms according to the number of relationships in which he is the center of attraction. He may at the same time be the nucleus of one or more atoms and part of the "electronic system" of other atoms. These atomic structures may therefore overlap in a tremendous intricacy of interrelationship, each individual being at the same time part of numerous diverse structures, but in various degrees of completeness or intensity.[30]

As was to be expected, Moreno found that a knowledge of the societal energy patterns of his community enabled him to explain events which occurred, such as a series of runaways, and, what is more important, it also enabled him to organize the community so as to relieve tensions, and to dissipate certain types of maladjustment. He also suggests the applicability of these methods in the social as well as in the physical planning of other than institutional communities, for example, such as those conducted by the Resettlement Administration.[31]

It should be recognized at the outset that in all of these studies of the interpersonal structure of communities, we are not under the delusion that we are charting *all* the rela-

[28] The nucleus is here defined merely as the center of converging societal currents.

[29] Atom is here defined as in other sciences, namely, as the smallest independent behavior unit with which it is possible to deal in the existing stage of development of knowledge and technology.

[30] See F. S. Chapin, *op. cit.*, ch. 16; "Measuring the Volume of Social Stimuli," *Soc. Forces*, 4:479-95, 1926; G. A. Lundberg, *Foundations of Sociology*, (Macmillan, 1939), Chap. IX.

[31] See the references to the articles by C. P. Loomis at the end of this chapter.

tionships that are operating. To secure the basic data necessary for a *complete* account of the interpersonal relationships in a community, protracted study through participant observer and probably psychoanalytic and other techniques would be necessary. We may assume that in every community there are, in addition to the avowed attractions and repulsions, a large number of illicit, illegal, and so-called "subconscious" likes and dislikes. Information about these is not easily secured through questionnaires or formal interviews. On the other hand, if we allege the existence of such relationships, the scientist must be prepared to *indicate objectively the behavior* (verbal and/or other) on the basis of which he declares the relationships (including the "subconscious") to exist. It may be sufficient for the artist to declare that he knows by intuition, introspection, or understanding, that such unconscious relationships exist. But this will hardly satisfy a scientist, however valuable these hunches may be as sources of hypotheses.

In short, we cannot take cognizance in science of relationships and social structures which are alleged to be so intangible that we cannot observe them by *any* verifiable technique. Such relationships may be *postulated* as *hypotheses,* but they should not be mistaken for scientific *knowledge.* We need not claim, on the other hand, that the relationships which can be observed objectively and verifiably at present are the only ones that will ever be subject to such observation. We may assume, on the contrary, that increasingly subtle social relationships will be observed as our techniques improve. In the meantime, we are justified in proceeding with the systematic analysis of *those* social relationships for which the data are readily available through questionnaires or brief interviews.

The type of results which are possible through mere interviewing is illustrated in Figure 13. The relationships indicated in this diagram were discovered simply through a casual question at the end of an interview concerned primarily with material and cultural possessions involved in the measurement of socio-economic status as described in the preceding chapter. The question was: "Who are your best friends in this community, *i.e.,* with whom do you most like to visit

'socially'?" The patterns revealed by the replies received are depicted in Figure 13.

In the whole community (population about 1000) eight such constellations, differing widely in pattern, were found.[32] These were analyzed from a number of viewpoints. Only a few of the more general results can be given here. (1) The groupings are correlated to some extent with socio-economic status. Groups of high socio-economic status were relatively homogeneous in this respect as compared with groups of low status. (2) The average number of choices *made* does not vary significantly with socio-economic status, but the average number of choices *received* varies uniformly and directly with the socio-economic score of the receiver. (3) No clear occupational bases for the constellations are evident, as every constellation has a great variety of occupations represented in it. (4) Family relationship is not prominent in these groups. (5) Geographic factors are not prominently correlated with the groupings in this village. (6) Membership in a common church is one of the factors most clearly associated with these friendship groupings. (7) *Ex officio* friends of everybody, such as ministers of the gospel, do not figure prominently in these groupings. Politicians, physicians, and officials were reluctant to mention special friends, but received many choices.

D. SOCIOLOGICAL INDICES

While charts and maps of the type illustrated in this chapter may be of great value in depicting a particular concrete situation, if we are to use these cases for scientific generalizations, we must develop methods of classifying, comparing, and summarizing them in terms of the frequency and the probability of their occurrence. It is in this connection that we become interested in the development of quantitative indices of the degree to which the observed relationships tend to appear and vary under different conditions. We saw above how some such indices have been developed in the case of ecological maps depicting geographic relationships. The de-

[32] For a full presentation and analysis of the data, see G. A. Lundberg and Mary Steele, "Social Attraction Patterns in a Village," *Sociometry*, 1:375-419, 1938.

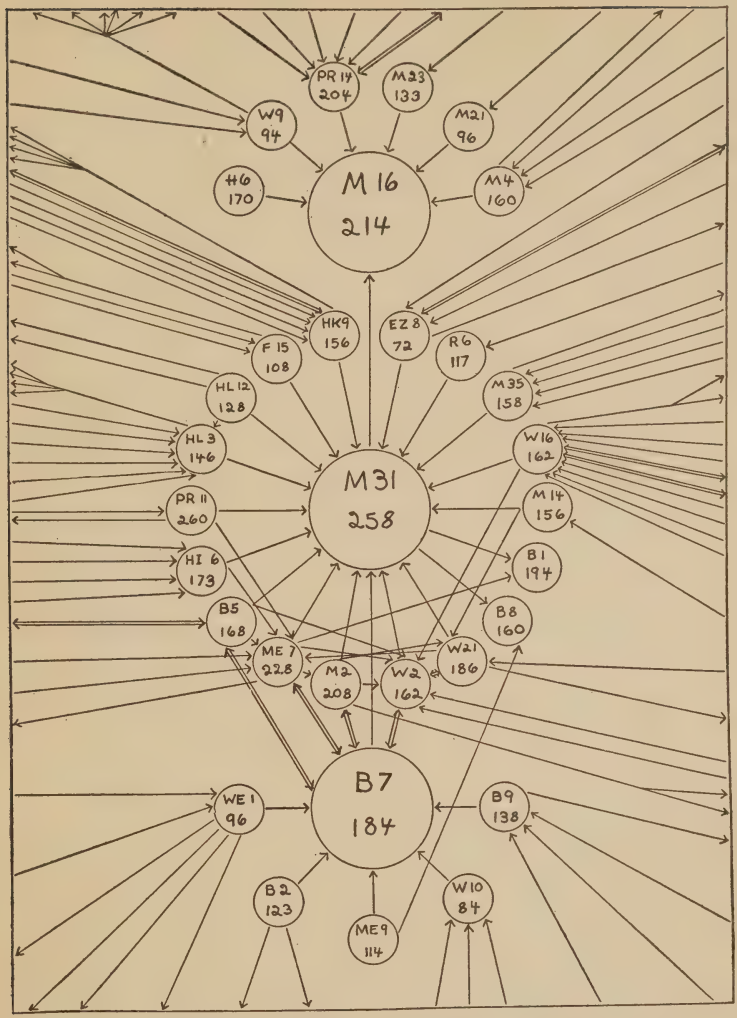


FIGURE 13. Friendship constellation in a village. The most important "friendship" constellation in the village, centering on the matriarch of the leading family with her principal satellites, a physician-politican (M 16) and a banker's wife (B 7). Each person is represented by a circle. The letter and the first number in the circle is the code symbol of the person. The second number is that person's score of socio-economic status on the Chapin Scale. Each arrow represents a choice made or received according to the direction of the arrow. Mutual choices are represented by double-headed arrows. (On account of charting difficulties the following lines are omitted from this chart: Me 7 and M 2 to Pr 14; We 1 and W 16 to Hl 3.)

gree to which many kinds of social phenomena are associated with these geographic patterns is thus indicated. What is more, for scientific purposes, the degree to which a relationship is observed to hold in one case can be compared and generalized, by means of these indices, for human communities in general, or at least for communities in fairly homogeneous cultures. If in addition to the facts regarding the frequency and direction of choices depicted in our social structure charts, we could develop also indices of the intensity of these attractions (not to mention repulsions), we could determine with greater accuracy how general the relationships are. The accurate determination of how these relationships vary with other factors also awaits the development of reliable indices of the relationships and the factors which we think are relevant in producing observed effects. We conclude, therefore, with a brief consideration of some such proposed indices. Some of them are frankly heuristic, and only further research can determine to what extent the phenomena to which these indices refer are of importance in sociological analysis.

1. Suppose it is desired to compare different groups on the basis of the amount of interaction taking place between the members of each group, *as measured by any observed behavior.* In Figure 13, for example, the number of choices of friends is taken as one indication of interaction. An index of interaction representing the percentage of total possible choices actually exercised between two or more people may be computed by the following formula:

$$\text{Index of interaction (expressed as a percentage)} = \frac{\text{Total choices made} \times 100}{N^2 - N}$$

The maximum number of choices possible among a given number of individuals is the square of the number of individuals in the group, minus this number, or $N^2 - N$. Thus, the maximum number of choices possible between two individuals is 2, among three individuals 6, among four individuals 12, etc. If, in the last case, the number of choices actually made is 9, the index of interaction is $\frac{9 \times 100}{12} = 75\%$.

2. The above index measures the degree of interactivity

(as indicated by the criterion of *choices*) among the members of a group. It tells nothing about the forces drawing that group apart or outward as balanced against the forces that hold it together, namely, the forces of intra-activity plus the forces pressing inward upon it from the outside. This latter phenomenon (accepting again expressed choices as an indication of it) may be measured by a compound index of cohesion or measure of centripetal-centrifugal tendency computed as follows:

$$\frac{\dfrac{\text{Number of choices made among members of an in-group plus the number of choices received by them from outside}\ [33]}{\text{Number of people in the in-group}}}{\dfrac{\text{Number of choices made outward by the in-group}}{\text{Number in the out-group receiving choices from the in-group}}}$$

Or,

$$\frac{\text{Per capita choices within a group plus per capita choices received from outside}}{\text{Per capita choices received by an out-group from the in-group}}$$

Or,

$$\frac{\dfrac{Ca + Ci}{I}}{\dfrac{Co}{O}} = \frac{O\,(Ca + Ci)}{ICo}$$

[33] This formula assumes that incoming choices are a cohesive or binding force in their operation upon the group receiving them. While this is doubtless true in some studies, it need not always be true. For example, if the in-group should consist of highly marriageable men and the choices should come from marriageable women, the incoming choices might readily be a disruptive force as far as the in-group is concerned and should therefore be subtracted rather than added in this numerator. I am indebted to Dr. Paul Lazarsfeld for this suggestion. He also points out that the formula as it stands gives great weight to the *concentration* of the in-group's outgoing choices as represented in the last term of the formula (last denominator), and suggests that this factor could be given a more moderate and proportional weight by changing the last form of the formula to read

$$\frac{\dfrac{Ca}{I} + \dfrac{Ci}{I}}{\dfrac{Co}{O} + \dfrac{Ci}{I}}$$

For other proposed indices of the phenomenon here under discussion together with some results of their application to actual groups, see J. L. Moreno and Helen K. Jennings, "Statistics of Social Configurations," *Sociometry*, 1:342-74, 1938. See especially pp. 367-74.

In which

I = Number in the in-group
O = Number in out-group receiving choices from the in-group
Ci = Number of choices coming in
Co = Number of choices going out
Ca = Number of choices among members of the in-group

Take, for example, an in-group of six members who make a total of 17 choices among themselves and receive 12 from the outside, a total of 29, a per capita of 4.83 choices. This is the numerator, a measure of cohesion *within* the group and pressing on it from the outside. This in-group makes six choices toward an out-group consisting of six persons, thus scattering its outgoing choices very thinly, namely, 1.0 choices per individual in the receiving group. The index of cohesion is accordingly 4.83 (4.83 ÷ 1.00). If these six outgoing choices had been concentrated on, say, one individual instead of six, the denominator in our fraction would have been 6.0 instead of 1.00, and the index of cohesion of the central group would be $\frac{4.83}{6.0}$, or .80. This would indicate (whenever the index of cohesion is less than unity) that the six members of the inner group actually are more strongly drawn toward the hypothetical individual assumed than to the members of the in-group, and should therefore be considered as satellites of the hypothetical individual. This index also indicates whether the peripheral group is more closely drawn toward the center of a group, such as is represented in Figure 13, or toward the margin. It is a sort of indicator of a group boundary in that it indicates at what point more forces draw a group toward one center rather than to another.

We are not concerned in the above indices with the question of whether expressed choices of friends as secured in a particular investigation is a desirable or a reliable indicator of interaction or cohesion. The proposed index merely provides a method of comparing different groups, whatever behavior is adopted as indicative of group cohesion or interaction. It is not contended that choices of friends is an adequate, the only, or the best indicator of group interaction or cohesion. That is a matter for research to determine. If these indices can be shown to vary significantly and consist-

ently with other sociologically important group differences of the type that are commonly designated as characterizing "closely knit," "ethnocentric," "self-centered," "self-sufficient," "homogeneous" in-groups, then the proposed indices may be used as measures of the degree to which groups have these characteristics. All indices are merely short and facile devices of this kind. In any event, there is no mystery about these indices. They specify exactly what they take into consideration, and how they take the items into consideration, as contrasted with such words as cohesion and interaction.

3. One of the principal problems that arises in the analysis of sociograms showing the distribution of choices, or other evidences of positive or negative interaction, is to determine to what extent the observed configuration differs from the interactions that may occur on a purely chance basis. For example, Moreno and Jennings[34] report the following experiment:

Twenty-six girls are allowed three choices each as to what members of the group they would prefer for dining-table companions. These data are secured for seven such groups. The average number of people receiving 0, 1, 2, 3 . . . 11 choices is computed. The question now arises how these results differ from what would be secured if, instead of making deliberate choice, each girl had drawn at random three names from an urn containing all the names but her own. In either case we would expect a certain number of people not to be chosen at all, a certain number of reciprocal choices, a certain number of mutual choices, etc. Only to the extent that the deliberate choices vary significantly from the configurations produced by chance do we have evidence of social attractions and repulsions in the group. The actual results secured in this experiment by (*a*) deliberate choice, (*b*) random drawing of names and (*c*) theoretical expectancy computed by formula[35] were as follows:

[34] J. L. Moreno and Helen H. Jennings, "Statistics of Social Configurations," *Sociometry*, 1:342-74, 1938.

[35] The computations shown in the table were made by Dr. Paul Lazarsfeld. (For full details see J. L. Moreno and Helen H. Jennings, *op. cit.*, pp. 349-51.) Under the conditions stated, the probability of a certain person being chosen by any other person is $p = \frac{3}{25}$; the probability of not being chosen is

TABLE VI

Frequency of Types of Relationships in a Group

Relationships	*Average Number of Each Type of Relationship*		
	Deliberate Choice	*Random Choice Experiment*	*Theoretical Chance Expectancy*
Isolated	5.0	1.4	1.1
Unreciprocated	51.1	69.4	68.64
Mutual	13.4	4.3	4.68
Chain relations	2.3	0.9	not computed
Closed structures (triangles, etc.)	1.0	0	not computed
Leader structures	6.6	4.7	not computed

It will be observed that when the girls choose their table companions deliberately, about four times as many people are not chosen at all as would be the case if chance alone were operating. Likewise, deliberate choice results in about three times as many mutual choices as chance alone would produce. Conversely, and as might be expected from the above results, the average number of unreciprocated choices is notably less under deliberate choice than under conditions of chance. The extent to which the observed configurations thus differ significantly from chance configurations indicates the degree to which social attractions and repulsions are genuinely operative.

4. A *sociation index* (S) for groups and individuals is sug-

---

$q = 1 - p = \frac{22}{25}$; the probable number of persons who, by mere chance, would be picked by their fellows *not at all, once, twice,* etc., is given by carrying out the binomial expansion

$$\left(\frac{22}{25} + \frac{3}{25}\right)^{25}$$

and multiplying by the number of persons. The general formula for n persons, each permitted a choices, is $(p + q)^{n-1}$ in which $p = \frac{a}{n - 1}$ and $q = 1 - p$.

The probability that any child is mutually chosen by any other child is given by the general formula $m = \frac{n(n - 1) \times p^2}{2}$.

The probable frequency of unreciprocated choices by mere chance is given by the formula $u = n(n - 1)pq$. See C. H. Forsyth, *An Introduction to Mathematical Analysis of Statistics,* (Wiley, 1924), Chap. 9.

gested by Zeleny[36] as follows: Each member of a group is asked to mark the names of every other member with the words "like," "dislike," "indifferent," according to the attitude felt toward working in class groups with the other members of the class. (The method is, of course, equally usable with regard to any other attitudes or groups.) In addition, the members are asked to indicate their first five (or other number of) choices of co-workers and also their *last* five (or most objectionable) choices. The *sociation index* is the average of the units of intensity of all the interpersonal attitudes in a group ($\bar{I}$), plus or minus the average deviation of the intensities of the attitudes from $\bar{I}$. This deviation is represented in the formula by (D). That is,

$$S \text{ (sociation index)} = \bar{I} \pm D.$$

Intensity of attitude is in the present illustrations assigned only two gradations: each acceptance (or rejection) received by a person is assigned one half unit of intensity (+ 0.5 or − 0.5).[37] If in addition to an acceptance (+ 0.5) a person receives also a "choice" (*i.e.*, is not only *acceptable* but is indicated by someone as among his first five choices in point of desirability) he is credited with an additional half unit of attraction (+ 0.5 + 0.5 = 1.00). Likewise, if a person receives a rejection (− 0.5) and also a "last choice," (− 0.5) this gives an (I) of − 1.00. The sum of these intensities

[36] L. D. Zeleny, "Measurement of Sociation," *Amer. Sociol. Rev.*, 6:173-88, 1941. For an actual community study dealing with the problem here under consideration, see C. P. Loomis and D. M. Davidson, Jr., "Measurement of the Dissolution of In-Groups in the Integration of a Rural Resettlement Project," *Sociometry*, 2:84-98, 1939. For some important considerations regarding the statistics of configurations and the normal distribution curve, see Helen H. Jennings, "Quantitative Aspects of Tele Relationships in a Community," *Sociometry*, 2:93-100, 1939. For a more complete and general discussion of the theory of measurements of this kind, see S. C. Dodd, "A Tension Theory of Societal Action," *Amer. Sociol. Rev.*, 4:56-77, 1939. See also the same author's *Dimensions of Society*, (Macmillan, 1941).

[37] "No attempt is made here to distinguish between the intensity of first, second and third choices; further research may or may not make these distinctions possible. This means, too, that it is not known whether a second half unit of intensity (a choice) is the equivalent of twice the first half unit of intensity, the square of it, or some other multiple. For the present, a linear relationship is assumed; and because of the satisfactory results so far obtained this relationship is assumed to be suitable for practical use at least.

"Note also that the unit of magnitude of intensity of an attitude is so chosen that the *sociation index*, $S = \bar{I} \pm D$, may have unity for its maximum value and thus vary in magnitude from $(+1.00 \pm 0.00)$ to -1.00 ± 0.00." (L. D. Zeleny, *op. cit.*, p. 174.)

(ΣI) divided by the total number of possible attitudes (n) in the group gives the average of all the units of intensity

$$(\bar{I}),\ i.e.,\ \frac{\Sigma I}{n} = \bar{I}.$$

The average deviation of $\bar{I}$ is computed in the usual manner, namely,

$$\frac{\Sigma d}{n} = \frac{\Sigma\,(\bar{I} \sim I)}{N(N-1)} = \text{average deviation}$$

in which N is the number of persons in the group. In summary,

$$S = \bar{I} \pm D = \frac{\Sigma\, I}{N(N-1)} \pm \frac{\Sigma\,(\bar{I} \sim I)}{N(N-1)} \text{ units.}$$

Suppose, for example, that in a group of three people, the attitudes expressed are as indicated in the numerator of the equation below. The sociation index would be computed as follows:

$$\bar{I} = \frac{\Sigma\, I}{N(N-1)} = \frac{+1 + 0.5 + 0 + (-0.5) + 0 + 0.5}{3(3-1)} = +\frac{1.5}{6}$$

$$= 0.25 \text{ units.}$$

$$D = \frac{\Sigma\,(\bar{I} \sim I)}{N(N-1} = \frac{.75 + .25 + .25 + .75 + .25 + .25}{3 \times 2} = \frac{2.50}{6}$$

$$= 0.42 \text{ units.}$$

Hence, $S = \bar{I} \pm D = +0.25 \pm 0.42$.[38]

5. Zeleny goes on to show how the sociation index becomes a *social status index* for an individual by a simple modification:

$$SS = \frac{\Sigma\, I}{n} \pm \frac{\Sigma\, d}{n} = \frac{\Sigma\, I}{N-1} \pm \frac{\Sigma\, \bar{I} \sim I}{N-1} \text{ units.}$$

This is the same as the sociation formula, except that N − 1 is substituted for N(N − 1) since we are considering here only the attitudes of other members of a group to a certain

[38] Note that the Average Deviation, although written after the calculated mean in the manner employed for indicating certain limits of observational error in science generally, here merely indicates the average amount of deviation of each intensity measure from the mean of all of these measures. The standard deviation may of course be used instead of the average deviation by substituting

$$\sqrt{\frac{\Sigma\, d^2}{n}} \text{ for } \frac{\Sigma\, d}{n}.$$

individual, and neglecting this individual's attitudes toward other members. To illustrate, a person receives 2 half units of attraction from person B and one half unit of repulsion from person C. Then his social status in that group as measured in terms of the social units is

$$SS = \bar{I} \pm D = \frac{\Sigma I}{N-1} \pm \frac{\Sigma(\bar{I}) \sim I}{N-1} = \frac{+1 + (-0.5)}{3-1} \pm \frac{0.75 + 0.75}{2}$$

$$= + 0.25 \pm 0.75 \text{ units.}$$

SS's may range from $+ 1.00 \pm 0.00$ to $- 1.00 \pm 0.00$.[39]

6. Finally, the sociation index becomes a *social adjustment index* by the following modification:

$$SA = \frac{\Sigma I}{n} \pm \frac{\Sigma d}{n} = \frac{\Sigma I}{2(N-1)} \pm \frac{\Sigma(\bar{I} \sim I)}{2(N-1)} \text{ units.}$$

Social Adjustment is here defined as the average intensity of *mutual* attractions and repulsions between a person and each of his associates in a group, plus or minus the average deviation of the measures of intensity of attitude around the mean. We are here concerned with the attitudes of the individual in question *toward* his associates as well as with the attitudes of the others toward him. Hence the base for the index becomes $2(N - 1)$ instead of $N(N - 1)$, as in the index of sociation, and $(N - 1)$ as in the social status index. $(N - 1)$ represents the number of attitudes a group may hold toward a person; since in a mutual relationship the person concerned may hold an equal number of attitudes toward the members of the group, $2(N - 1)$ represents this number.

Zeleny has shown how these indices can be used in social experimentation by rearranging groups and observing the effect as measured by the indices as well as by other criteria tending to validate the indices. Others have used similar techniques in therapeutic measures for maladjusted groups.[40]

[39] The status of a person in a community might be the average of his statuses in all the groups to which he belongs. The present illustration does not deal with this problem, however. See L. D. Zeleny, "Measurement of Social Status," *Amer. Jour. Sociol.*, 45:576-82, 1940. On the measurement of morale, see also D. C. Miller, "Morale of College Trained Adults," *Amer. Sociol. Rev.*, 5:880-89, 1940. Also E. A. Rundquist and R. F. Sletto, *Personality in the Depression*, (Univ. of Minnesota Press, 1936).

[40] See Leona M. Kerstetter, and J. Sargent, "Reassignment Therapy in the Classroom," *Sociometry*, 3:293-306, 1940. Also Joan H. Criswell, *A Sociometric Study of Race Cleavages in the Classroom*, Archives of Psych., New York, 1939.

It should be emphasized again that these indices do not pretend to designate or measure *all* of the aspects of the phenomena denoted or connoted by such words as sociation, interaction, status, or adjustment. On the other hand, the indices designate precisely what they do take into consideration, and represent the kind of tools of analytical thinking which every science seeks. Whether and to what extent these indices prove useful in sociological analysis and social control is a question to be settled by future research of the type mentioned above.

7. *An index of social turnover or metabolism.*—The term social metabolism has been used to designate the rate of shift in the personnel composition of groups. It is a general term for what in a special case in economics is called labor "turnover." E. Y. Hartshorne,[41] after reviewing the various indices that have been proposed in this field, suggests the following *stability index*:

$$\text{stability index} = \frac{2\ R}{G_0 + G_1} \times 100.$$

In which

G_0 = the size of a group at the time subsequent to which we wish to measure change.

G_1 = the same group, (nominally, at least) after change during a lapse of time, during which some or all of the original members have been lost and new ones have been added.

R = the residue who have survived the period of change.

That is, if a university has a faculty of 615 at one time, loses 304 members, leaving 311, and adds 14 new ones, thus having 325 at the end of the period, its stability index for the period is:

$$\frac{2 \times 311}{615 + 325} \times 100 = \frac{622}{940} \times 100 = 66.1\%.$$

It will be observed that 100% stability (or 0% turnover) would be shown by this index only if R, G_0, and G_1 were the same. Likewise 0% stability (or 100% turnover) would

[41] E. Y. Hartshorne, "Metabolism Indices and the Annexation of Austria: A Note on Method," *Amer. Jour. Sociol.*, 45:899-917, 1940.

be shown only when R = zero, *i.e.,* when all members are replaced.

As in the case of all indices, this index does not pretend to describe all significant aspects of the metabolic or turnover process. The operations involved in its computation show exactly what aspects it does take into consideration, and it should not be expected to measure anything but these aspects, *and whatever aspects may be shown to correlate to a stated degree with those measured by the index.* If this principle were kept more generally in mind both in inferential reasoning from indices and in criticisms of social indices, a great deal of futile controversy would be avoided.

E. SUMMARY AND CONCLUSION

The discovery of systematic relationships among the multitudes of events that occur in, among, and around us is, as we have seen, the central problem of all science. This quest has been immeasurably advanced through the use of written symbols by means of which the uniformities of recurring events have become demonstrable. By means of symbols, also, enormous areas of data are brought together in sufficiently compact form to reveal the interrelationships between the parts. For example, a city map the size of this book page reveals at a glance geographic, ecological, and demographic relationships which could not be apprehended in a long time by a casual stroller. We considered, accordingly, in the first part of the chapter, the cartographic devices which have been found useful in depicting social conditions and relationships, against a background of geographic space.

We found in the second place, that cartographic techniques were also useful in depicting purely social relationships without any reference to geographic space. Charts of social organization and functioning are almost as common as geographic maps. These social structures and functions are conceived to exist in *social space,* as when we speak, for example, of the president being *above* the members of his cabinet, the city manager being *under* the city council, etc. These spatial designations indicate certain *social* relationships and functionings, and in this way are as useful as spatial designations indicating which part of a city, or which army, is *above*

another in a purely geographic sense. In the same way, we saw that social structure, status, interpersonal relations, and many so-called "abstract" and "intangible" relationships may also be depicted objectively by geometric, topological, and sociometric diagrams and charts.

Finally, we showed that as we wish to refine and sharpen the content of the patterns found by the above methods, and to generalize our observations for scientific purposes, we need indices denoting the extent, degree, and particular aspects of the relationships, roughly indicated in the charts. That is, a map showing the relative positions of two cities is of importance if we want to build a railroad between them. If, in addition, the map shows in standard units the distance in miles and the exact location and elevation of hills and valleys, this adds vastly to the usefulness of the map. Likewise, a simple diagram of certain relationships among members of a particular group may be useful for some purposes. But if we wish to compare this group with others, and especially if we want to arrive at generalizations about the observed relationships, we need symbolic devices designating exactly what we are comparing, such as the kind and intensity, of the relationships. Indices are designed to supply this need.

Indices—rates, coefficients, and formulas of all kinds—are especially designed to symbolize *briefly* and *precisely* important relationships among phenomena. Accordingly, all sciences strive to develop reliable indices of the phenomena with which they deal. A vast variety of indices for specific uses have, therefore, been developed in every field, including the social sciences. The practical demands of business and education have greatly stimulated these developments in economics and psychology. We have in previous chapters considered at some length the attempts to develop similar indices of important social relationships. The additional illustrations given in the present chapter, together with those discussed in other chapters, do not, of course, represent a complete list of social and sociological indices now proposed and in use.[42] We have aimed rather to illustrate recent at-

[42] The most comprehensive proposals for indices of a very large variety of sociological phenomena will be found in S. C. Dodd, *Dimensions of Society*, (Macmillan, 1941). See especially Chap. 10.

tempts to develop indices of phenomena hitherto considered largely unapproachable by more precise and objective methods.

We have warned repeatedly against taking a mystical view of indices, and of using them blindly. Mysticism about indices is found perhaps chiefly among those who assume that because an index is called by some familiar word, it necessarily measures all of the things which that word means in its vague folk usage. The requirements of definiteness and precision in an index probably preclude the possibility that its denotation should ever be the same as the denotation and connotation of a particular folk word or phrase. When this fact is generally realized, we shall cease to expect indices to give us the *same* meanings as folk terms. We shall strive instead to develop indices that will define sociological concepts in terms of the operations involved. The usefulness of these definitions will be demonstrated, if at all, by the results they yield; by the extent to which they enable us to reach significant, verifiable generalizations. In this way indices, like other tools, although themselves the products of analytical thought, will become in turn the instruments of more refined analysis.

F. SUGGESTIONS FOR FURTHER STUDY

1. One of the best examples of the use of a variety of methods in human ecology is Calvin F. Schmid, *Social Saga of Two Cities,* (Minneapolis Council of Social Agencies, 1937). A good brief treatment of these methods will be found in the same author's "The Ecological Method in Social Research," (Chap. 14 in Pauline Young, *Scientific Social Surveys and Social Research,* Prentice-Hall, 1939).

2. For a remarkable but somewhat technical example of the usefulness of ecological data in the verification of sociological hypotheses, see S. A. Stouffer, "Intervening Opportunities: A Theory Relating Mobility and Distance," *Amer. Sociol. Rev.,* 5:845-67, 1940.

3. The following sources, in addition to those reviewed in the text, present illuminating examples of both methods and

results of sociometric analysis of social interaction in human groups:

(*a*) J. L. Moreno, *Who Shall Survive? A New Approach to the Problem of Inter-human Relations* (Nervous and Mental Disease Publishing Co., Washington, D. C., 1934).

(*b*) F. J. Roethlisberger and William J. Dickson, *Management and the Worker,* (Harvard University Press, 1939). (This excellent monograph employs a variety of methods. See especially Chaps. 20-22.)

(*c*) C. P. Loomis and D. M. Davidson, Jr., "Measurement of the Dissolution of In-Groups in the Integration of a Rural Resettlement Project," *Sociometry,* 2:84-94, 1939.

(*d*) C. P. Loomis and D. M. Davidson, Jr., "Sociometrics in the Study of New Rural Communities," *Sociometry,* 2:56-76, 1939.

(*e*) C. P. Loomis, "Informal Groupings in a Spanish-American Village," *Sociometry,* 4:36-51, 1941.

(*f*) Joan H. Criswell, *A Sociometric Study of Race Cleavages in the Classroom* (Archives of Psychol., No. 235, New York, 1939).

(*g*) W. I. Newstetter, M. J. Feldstein, and T. M. Newcomb, *Group Adjustment, A Study in Experimental Sociology,* (Western Reserve University, Cleveland, 1938).

(*h*) Helen H. Jennings, "Structure of Leadership—Development and Sphere of Influence," *Sociometry,* 1:99-143, 1937.

(*i*) Mary L. Northway, "A Method of Depicting Social Relationships," *Sociometry,* 3:144-50, 1940.

(*j*) D. C. Miller, "Evaluative Research in Group Discussion," *Sociol. and Social Research,* 25:213-25, 1941.

(*k*) F. S. Chapin, *Contemporary American Institutions,* (Harper, 1935), Chaps. 2-6, 16-18.

4. For an excellent summary of sociometric work, see F. S. Chapin, "Trends in Sociometrics and Critique," *Sociometry,* 3:245-62, 1940. A general annotated bibliography of sociometric techniques of group analysis has been compiled by

J. G. Franz, "Survey of Sociometric Techniques with an Annotated Bibliography," *Sociometry,* 2:76-90, 1939.

5. Extensive development of theoretical and quantitative aspects of sociometrics will be found in S. C. Dodd, *Dimensions of Society,* (Macmillan, 1941). For briefer treatment of particular aspects, see his "A Tension Theory of Social Action," *Amer. Sociol. Rev.,* 4:56-77, 1939; his "The Interrelation Matrix," *Sociometry,* 3:91-101, 1940; Helen H. Jennings, "Quantitative Aspects of Tele Relationships in a Community," *Sociometry,* 2:93-100, 1939; E. D. Chapple, "'Personality' Differences as Described by Invariant Properties of Individuals in Interaction," *Proceedings of the National Academy of Sciences,* 26:10-16, 1940. A more complete account of this technically difficult study will be found in the same author's monograph (in collaboration with C. M. Arensberg), *Measuring Human Relations: An Introduction to the Study of the Interaction of Individuals.* Genetic Psychology Monographs, 22:3-147, 1940.

6. On topological methods in psychology and sociology, see J. F. Brown, *Psychology and the Social Order,* (McGraw-Hill, 1936), Chaps. 3 and 4 and Appendices A and B. For more extensive treatment, see K. Lewin, *The Conceptual Representation and the Measurement of Psychological Forces,* (Duke University Press, 1938).

7. On social indices and their uses, see P. F. Lazarsfeld, "Interchangeability of Indices in the Measurement of Economic Influences," *Jour. of Applied Psychol.,* 23:33-45, 1939. See also B. Buell and R. Robinson, "A Composite Rate of Social Breakdown," *Amer. Jour. Sociol.,* 45:887-98, 1940. Also F. S. Chapin, *Contemporary American Institutions,* (Harper, 1935), Chap. 20, "A Measurement of Personal Intimacy." An excellent discussion of the function and importance of indices in the social sciences with illustrations from the work of Durkheim will be found in Harry Alpert, *Emile Durkheim and His Sociology,* (Columbia University Press, 1939), pp. 119-27.

8. For a very important type of social measurement not treated in this book see G. K. Zipf, *National Unity and Disunity,* (Principia Press, Bloomington, Ind., 1941). See also F. H. Allport, and R. S. Solomon, "Lengths of Conversations:

A Conformity Situation Analyzed by the Telic Continuum and the J-Curve Hypothesis," *Jour. of Abnorm. and Soc. Psychol.,* 34:419-64, 1939. Also A. Jenness, "A Critique of the Telic Continuum and the J-Curve Hypothesis," *Psychol. Bull.,* 36:531-32, 1939. Also F. H. Allport, "The J-Curve Hypothesis of Conforming Behavior," *Jour. of Social Psychol.,* 5:141-83, 1934.

CHAPTER XI

FIELD WORK: THE INTERVIEW AND THE SOCIAL SURVEY

Let us assume, however, that we are now satisfied that we know where to find, and have set up a sample truly representative of the country as a whole, and that we have found words and phrases which clearly and without bias convey our meaning. We are still confronted with the enormous job of getting the actual field work done. . . No matter how skillfully drawn the questionnaire is, it can fall down if it is sent to fifty deserving college juniors [as interviewers] who have no feeling of the great importance of their part of the work and who have not been thoroughly trained. . . We are often asked what kind of people make the best interviewers and find it very difficult to "type" them. They must all be honest; they must all have personalities sufficiently attractive to let them sell respondents on the idea of giving an interview; they must realize the importance of letting no inflections creep into their voice which might influence answers; and of course, they must be possessed of a good strong pair of legs! We are materially aided, however, in our efforts to secure competent people by the fact that the work itself is extremely interesting and appeals to intellectually curious and intellectually honest people. Until last year one of our interviewers was one of Hollywood's foremost character actors who stated that our work gave him a better opportunity to study character types than anything else he did. Another was the vice-president of a nationally-known manufacturing concern who liked to spend all of his week-ends interviewing because of his genuine interest in sociology. — *Elmo Roper.*[1]

A. INTRODUCTION

The major theme of this book has been the need for the accurate observation and recording of first hand data and their reduction to valid generalizations. We have treated (1) the importance of definite units and objective terminology (Chapter III); (2) the determination of the scope and purpose of our investigation; that is, the formulation of a working hypothesis and the selection of a representative sample (Chapters IV and V); and (3) the construction of mechanical devices in the form of schedules of various kinds to help us objectively to observe, record, and compare data gathered from original sources (Chapter VI to X). There remains the consideration of the actual field work.

We saw in Chapter VI that data may be secured either from documentary sources or from field work. Field work we classified as of two principal kinds: (1) direct observations

[1] "Sampling Public Opinion," *Jour. Amer. Statis. Asso.*, 35:329-30, 1940.

of any social phenomena, including verbal behavior; (2) the oral or written testimony of persons, when our interest is in the situations represented by the verbal responses rather than in the language behavior itself. When our interest is primarily in the verbal behavior *per se,* its observation belongs, of course, to the first category. It is recognized that the pursuit of documentary sources may involve field work, and to the extent that such is the case, the technique of field work would apply also to research from documentary sources. But since the technique of securing data from documentary sources has been relatively highly developed by the historians, and has been discussed in a previous chapter (Chapter IV), we shall confine our efforts in the present chapter to a consideration of the technique of gathering social data by direct observation of the behavior, verbal or otherwise, of persons and groups, and of their significant environmental conditions.

The best way to learn the technique of efficient field work is to participate in an actual field project under a capable director. The next best way is to study the printed reports of successful field investigations (see suggestions at the end of this chapter). The nature of field work will vary so greatly with the type of investigation as to make any detailed treatment more or less inapplicable to some specific situations. We shall not attempt, therefore, to do more than designate the principal problems likely to occur in all interviews. A more detailed consideration of field work must be left to specialized treatises on surveys and interviewing.

The technique of field work may be considered from three principal points of view: (1) the authority and backing of the investigator; (2) the organization of a field staff; and (3) the interview.

B. THE AUTHORITY AND BACKING OF THE FIELD WORKER

Since field work of any kind involves encroachment upon the time, premises, and information of other people, or upon the community as a whole, the first consideration is to justify such encroachment. When the field work is of an official character and is undertaken by the proper authorities in the fulfillment of their duties, this constitutes in itself the re-

quired justification and authority. Such is the case, for example, in the taking of the Federal or State Censuses, the inquiries of the local assessor, tax collectors, etc. But most field work primarily for research purposes is not of this type. There is, to be sure, a great deal of research or investigation carried on by official bodies — commissions, survey committees, etc. — which derives its authority from the government, and thus has the same justification and authority as the field work conducted by any other regularly constituted officials. But allowing for the research carried on by such bodies, there is still a great deal carried on by private individuals and private organizations which requires the cooperation of the community for its successful execution. In the physical sciences, for the most part, we seize our material, whether it be the lower animals or inanimate objects, and study it without consulting the material itself. But in social research, the passive if not the active participation of the individuals or the community to be studied is frequently essential. Herein lies, of course, one of the principal obstacles to social research as compared with research in the physical sciences.

The best method of securing the cooperation of the individuals or the community to be studied is to secure for the investigation the approval of leaders in the community. This might include such approval by the duly constituted authorities and thus make cooperation mandatory. More frequently, however, we must rely on the less overt social pressures — the recommendation of the priest, the church, the civic organizations — for a friendly attitude toward the research on the part of the persons touched by the inquiry. The first step, then, is to determine which persons' cooperation will be required in the proposed field work and, secondly, to secure for the project the approval of the organizations, officials, or leaders to whose judgment the persons concerned are accustomed to defer. The final cooperation of each individual concerned must, of course, be achieved by the technique of the interview, the letter, or whatever form of approach is employed. But the approval of the leaders of the community will materially facilitate the personal approach.

The approach and backing of important social and civic

agencies in the community is important not only from the standpoint of securing community cooperation, but the financial support of such agencies is frequently required in any field work which is at all comprehensive. While the number and types of organizations whose backing is desirable vary greatly with the scope and purpose of the projected field work, it is usually desirable to secure the backing of more than one agency, as not all people in a community may be closely associated with or pay much heed to any single agency. For field work involving the whole community, the backing of a committee representing all the important interests and activities in the community is most desirable. It insures an avenue of approach and a basis of appeal to everyone. In an age of high-power salesmanship, people are often properly wary of giving even apparently harmless information to a stranger, regardless of the explanation the inquirer may offer. The approval of an organization in which the informant has confidence becomes, therefore, an introduction for the field worker and an assurance to the informant that he may cooperate freely without the danger that the information which he gives will be used to his disadvantage. The publicity which can be secured for the investigation (when this is desirable), by having it sponsored by prominent representatives of important civic and social agencies, also prepares the public for the field work and greatly facilitates the task of the field worker. Whether financial assistance is sought or not, therefore, backing for any investigation involving field work should be secured from the organizations which command the confidence of the community. We are speaking here of "community" studies in which the cooperation of large numbers of officials and other members of the unit studied is essential. Private commercial studies, polls, etc., are, of course, successfully carried out without any official or group sanction.

C. THE ORGANIZATION OF A FIELD STAFF

Before the regular field work can be organized it is necessary for the director of the investigation, in addition to his general plans, to familiarize himself with the fields and the sources of information. Only in this way can he determine

the size and type of staff necessary as well as what persons, groups, or organizations are to be canvassed by his field workers. Who may be qualified field workers and the degree of organization and supervision required depends entirely upon the type of field work to be done. In many local surveys of general community conditions, volunteer field workers who are especially interested in certain aspects of community life, such as teachers, social workers, and students in special fields, may prove earnest and capable field workers. On the other hand, an inquiry of national scope, like the Federal Census, requires a large number of workers not previously trained, and in such cases the most careful organization and supervision becomes necessary, if the results are to be complete, comparable, and reliable. While the requirements of a staff will vary greatly according to the scope and nature of the inquiry, we may consider certain general principles which may be said to apply in greater or lesser degree to the selection, instruction, organization, and supervision of any field staff.

If the inquiry is of a specialized nature so as to require the services of specially trained field workers, these would be selected as specialized talent is selected for any other work. If the field work can be performed by persons without special training, provided they are carefully instructed, and if the necessities of the investigation require the utilization of such persons, the most important criterion of their desirability is that *they should be able and willing to follow instructions explicitly*. This does not mean that persons who have this mechanical ability will necessarily be desirable field workers. Most field work requires resourcefulness, initiative, and tact, as we shall see in our discussion of methods of interviewing. At the same time, persons who as field workers feel free to follow their own inclinations on what is to be achieved, are very difficult to manage, and their work is frequently worthless from the standpoint of the investigation as a whole. The essence of field work data, as of all other scientific data, is that they should be accurate and comparable. This end can be achieved only by a certain uniformity of procedure. When untrained workers must be employed, instructions as specific as possible, designed to cover every contingency that may

arise, should be supplied to each worker. Carefully prepared schedules with detailed instructions for filling them out are the best devices for securing uniformity and accuracy in the collection of field data.

Whenever possible, the employment of immature, untrained, and tactless field workers should be avoided in social investigation. In addition to the unreliability of the data collected by such persons, it is important not to create an unfavorable attitude on the part of the public toward the particular project and toward field investigation in general. As has been pointed out above, the success of much social research depends largely on the cooperative attitude of the groups studied. This cooperative attitude may be sacrificed by incompetent field workers, not only in a particular project, but for future projects as well. The practice of sending high-school and immature college students indiscriminately into the field to gather original data for term papers is to be deplored. It often becomes the basis of deep prejudice against all field workers in a community. Untrained field workers should not be employed without thorough instructions and close supervision.

The organization and supervision of a field staff, again, depends upon the scope of the investigation. The following account of the Federal Census of 1940 will give some idea of the organization required for this, the largest of all field inquiries: [2]

> More than ever before the 1940 Census has been a cooperative enterprise on a Nation-wide scale. In a very real sense the decennial census of the United States is a national undertaking, the Bureau of the Census being but the instrumentality for the census. Conferences, voluminous correspondence, and in some cases public trial, determined the questions to be asked and the tabulations to be made. Through the cooperation of newspapers, radio stations, motion pictures, trade associations, local citizens' committees, and other groups too numerous to mention, the Nation was informed about its census. And the Nation responded with a good will to give reports for more than 131,000,000 individuals, 36,000,000 dwelling units, 6,000,000 farms, 3,000,000 places of business, 180,000 factories, and 30,000 mines and oil wells.

[2] *Annual Report of the Secretary of Commerce*, 1940, pp. 37, 40-42, 59.

Every census has four distinct phases: (1) planning and preparation, (2) enumeration, (3) compilation of the data, and (4) preparation of final reports. . .

Space adequate for the 9000 employees required for the decennial census and other Bureau work at the peak during 1940 and 1941 was a pressing problem. In March 1940, the three largest divisions of the Bureau concerned with the decennial census, population, agriculture, and mechanical tabulation, were moved into the new Census Building, just completed. Although this modern building provides 396,000 square feet of space, the Bureau has had to rent additional space outside the Commerce Building for certain of its functions, and will utilize the new building for a double shift during the peak of the decennial census work. . . The following number of employees is approximately correct for the peak of field operations on the decennial census: total employees in field administrative offices (including supervisors, clerks, squad leaders, and interpreters), 7361; enumerators for population, agriculture, and housing, 101,916; enumerators for business and manufactures, 6396; and enumerators for irrigation and drainage, 261. . .

Employee recruiting and training. Nearly 130,000 persons will work in the field or in the Bureau's Washington office on some phase of the Sixteenth Decennial Census. Emphasis has been given both to the recruitment and the training of this army of temporary employees.

During September, October, and November, 1939, 180 candidates for the position of area manager were given an 8 weeks' training course in Washington. These classes were taught by census experts in their specific fields, and included instruction in administrative and personnel procedures, as well as in schedule forms and instructions. Objective tests on each census subject were given and the grades on these tests were correlated with the grades made on general intelligence tests, and together these grades formed part of the basis for the hiring and placement of these workers.

During December each area manager held a 1 week training program for his district supervisors using detailed training instructions prepared in the Washington office. The district supervisors in turn held from 1- to 4-day training classes for the census enumerators just prior to the beginning of their work. Again, objective tests were prepared for the selection and placement of enumerators.

Periodic correspondence instruction was given to the area and district supervisors from January through March. Regional training classes were held in nine cities for special instructions on population, agriculture, and housing during February and March.

Sound motion pictures were made in Washington and distributed to district supervisors to assist them in training enumera-

tors. This visual method of training was an innovation of the Sixteenth Decennial Census and was perhaps the largest visual training program ever given in so short a period. Instruction manuals, illustrative examples of completed schedules, and objective tests were prepared to assist the district supervisor in the training of enumerators.

An intensive program is being conducted for temporary office employees to train them in editing, coding, and machine tabulation.

Census publicity. A Division of Public Relations was organized in August 1939 under specific authority of Congress to plan and execute an educational campaign for enlisting Nation-wide cooperation in the decennial census.

Operating through various media of public expression — newspapers and periodicals, radio, motion pictures, local committee organizations, public officials, and the Bureau's field force — an intensive campaign was conducted in advance of and during the enumeration. More than 2000 cooperating local committees sponsored by chambers of commerce and State, county, and municipal officials assisted this program in their localities.

School officials, church leaders, civic groups, and fraternal organizations throughout the country also gave their assistance toward insuring the completeness of the census. Valuable contributions in services were received from county farm agents, home demonstration agents, and other leaders in the rural sections of the country. Other agencies of the Federal Government also cooperated in passing the word along through their local offices everywhere. The 40,000 postmasters throughout the Nation gave space on lobby bulletin boards to posters and other expressions of census information. . .

Editors of more than 10,000 daily and weekly newspapers, and of more than 1600 general magazines, business and professional periodicals, agricultural publications, and organs of trade associations and commercial organizations also cooperated in giving space to explanatory articles about the census as well as editorials urging the fullest cooperation on the part of their readers.

The Office of Education contributed much to the success of the radio program as did, of course, the various radio broadcasting chains, commercial program sponsors, and hundreds of individual radio stations. Similarly, the cooperation given by several thousand individual motion picture theater operators and by national news-reel companies is to be credited as a contribution of major importance.

Other agencies and groups, too numerous to detail in consideration of space limitations, contributed also — as, for instance, news services and press associations, Washington newspaper representatives, free-lance writers, volunteer speakers, library officials, heads of individual business organizations, and so on. The Bureau also

had the benefit of consultative service by technical experts on the various media of publicity. . .

Special provision was made for the enumeration of transients throughout the country. As the census enumeration is carried on over a period of several weeks, transients may be missed by the enumerator if they move during the enumeration period. In order to avoid this contingency, April 8 was set aside as the day when the usual places of residence of transients in all cities would be visited by enumerators. Special blanks were used to improve the enumeration of W.P.A., C.C.C., and N.Y.A. workers. Absent family schedules and nonresident schedules were used more extensively than in previous censuses, and were mailed directly to Washington for allocation to the proper enumeration districts. A card for new occupants was left in all vacant dwelling units to insure the enumeration of persons moving during the census period. The use of these supplemental forms was aimed directly at securing a more complete count of transient population than in previous censuses. It has been necessary to check these various forms against the names of the population schedules to avoid duplicate enumeration. It is evident that a good enumeration of transients has been secured.

Trial census. A special census of St. Joseph and Marshall Counties, Ind., was authorized by the Secretary of Commerce to be taken as of August 14, 1939. This census, a preview of the decennial census of population, proved to be very helpful in providing a testing ground for the schedules, auxiliary forms, instructions, and procedures planned for the decennial census and indicated some necessary changes in the schedules and procedures. Two innovations of the trial census were the use of objective tests as a means of testing and selecting the enumerators after a period of census training and the employment of squad leaders as supervisors of from 10 to 20 enumerators. Both of these innovations were adopted for the Sixteenth Decennial Census.

This trial census also provided real data for training office employees before the population schedules from the regular census were received. Preliminary editing and coding instructions, card forms, tabulations, and even table forms for the final census reports for 1940 were developed well in advance of any previous census on the basis of this substantial "census preview."

The precautions taken to secure a uniform type of field work in a project of much lesser scope is illustrated by the supervision of field work connected with the scoring of the health departments of eighty-six cities by the American Child Health Association discussed in Chapter IX.

Great pains were taken to insure a uniform interpretation of the schedule by the surveyors. Each surveyor, after all possible explanations before beginning, was accompanied on his first survey by the senior surveyor who was chiefly responsible for the preparation of the schedule. The meaning of the questions and the type of answers expected were thoroughly discussed. This was done to avoid misinterpretation and consequent lack of comparability of the answers. In addition to this precaution each report was subjected to critical inspection, and omission and nonspecific entries were taken up directly with the surveyors by correspondence. The value of this uniform supervision of each surveyor was unmistakable.[3]

Methods of handling cases of refusal and failure to contact informants are suggested in the following account:

Both preventive and corrective measures were adopted to meet the problems of refusals and of failures to contact families. Incomplete schedules and refusals are largely due to the approach used by the field agent, so the best preventive measures are directed at a careful training of agents and control of assignments. Calls at night to interview someone other than the person refusing, eliminating those agents who turn in too many refusals, shifting unsuccessful agents from the more difficult districts to the sections where families are more cooperative, reduced refusals. The agents in the present study were instructed to make at least two calls upon families before turning in the cards with partial or no information. Assignment supervisors were instructed to assign cases for a third visit to different and more skilled agents. Another control of the work of the agent was exercised through *check interviewing* which made it possible to discover agents who reported vacancies when families refused information. It was claimed by some supervisors that emphasis on speeding up production would increase the number of incomplete and unacceptable schedules. This was not found to be the case, however. In fact, the most productive agents also had the fewest refusals, and the most accurate schedules. A third important preventive measure consisted in securing favorable presentation of the investigation in the press.

Corrective methods in assuring the collection of the random sample as selected, consisted in measuring the biases introduced by refusals, incomplete schedules, and cases where families could not be contacted. The first analysis of any possible biases was made by finding the proportion of refusals and uncontacted cases in each district. Early reports indicated that a significant number

[3] American Child Health Association, *A Health Survey of* 86 *Cities,* (New York, 1925), p. 9.

of families in the wealthier neighborhoods were not at home or refused to give the information requested. During the hottest weeks of the summer months as many as 50 percent of the families in the better districts of the large cities had left the city and moved to vacation places. It was possible to revisit these families upon their return to the city in the fall after the early analysis had revealed the need for revisits.[4]

These examples will illustrate the types of organization and supervision of field work in cases where the data sought are specific and objective. Methods of organization and supervision must, of course, be modified to suit the particular investigation. Field work of a preliminary, exploratory, or "detective" nature cannot be as rigidly detailed in advance as in the above cases. But such exploratory field work must be regarded as preliminary to scientific surveying. It may produce clues and provide the basis for the systematic study of the situation. In such work the greatest possible freedom must be allowed the worker. But for the collection of data to form the basis of scientific generalization, the requirements of objectivity, uniformity, and comparability compel us to rely on more formal organization and supervision. Eternal vigilance in the form of constant checking is the price of accurate data in all scientific observation.

D. THE INTERVIEW

A complete account of the principles and practices of interviewing would involve a lengthy treatise on the technique of salesmanship, newspaper reporting, legal examination, including "third degree" methods, case work, and psychoanalysis, as well as an exhaustive consideration of the principles of psychology. Any one of these subjects is in itself a suitable topic for a volume. For our present purposes, however, we need consider only those aspects of the subject which relate to the more external technique of any field work involving interviewing. For a detailed consideration of the social and psychological processes involved, and the

[4] E. H. Schoenberg and M. Parten, "Methods and Problems of Sampling Presented by the Urban Study of Consumer Purchases," *Jour. Amer. Statis. Asso.*, 32:317-18, 1937.

specific technique to be employed, the reader must be referred to technical treatises on the subject.[5]

The interview may be of significance to social research in two principal ways. (1) It may be simply a method of securing data from a secondary source when our interest is mainly in the data that are conveyed to us by the informant instead of in his verbal behavior in conveying them. Thus the census taker or assessor inquires about certain objective facts as a short-cut method of securing them. As a rule he ignores the attendant behavior of the informant. When the information transmitted is our primary interest, the same principles of criticism that apply to any secondary source are applicable to the interview. (2) The interview may also be of importance to social research as a laboratory study of verbal behavior. In this case our primary interest is in the verbal responses, as such, to stimuli of various kinds. From such responses, the psychoanalyst, for example, makes inferences and draws conclusions.

In most interviews we are interested in both aspects of the process. Both types of interview involve to some degree essentially the same techniques of approach—securing appointment, establishing rapport, inspiring confidence, and releasing mental mechanisms. The researcher is interested in the objective data secured from the interview, such as income, the number of children, their ages, etc., and also in the personality of the informant—his attitudes: prejudices, likes, and dislikes—as revealed by his verbal behavior, including the subtle gestures that accompany it, such as facial expression, tone of voice, and so forth. As in any scientific observation, research by interview consists of selecting out and recording those aspects of the total situation in which we are interested.

1. *Selecting persons to be interviewed.*—The first problem which presents itself in research involving interviewing is the selection of persons to be interviewed. How definitely

[5] Some of these techniques are treated in Chaps. VII and VIII. Perhaps the best treatment of interviewing from the standpoint of the social worker is Pauline Young, *Interviewing in Social Work,* (McGraw-Hill, 1935). See also, E. S. Bogardus, *The New Social Research,* (Jesse R. Miller, Los Angeles, 1926), Chaps. IV, V.

such persons can be selected in advance depends upon how well-defined is the object of our inquiry and how completely the sources of such information are known. If we are interested in the wage scale of a certain industry, the responsible heads of the various establishments in this industry at once suggest themselves as the persons from whom to secure this information. The social worker seeking items of general information about a case consults neighbors, employers, relatives, teachers, and public officials. She may get the information from any, all, or none of these.[6] The problem of whom to interview, therefore, is a question of deciding who will yield most easily the desired information with the greatest reliability. Our familiarity with the field in which we are working will determine our ability to select such sources without employing the trial-and-error method.

In any case, all of the principles of valid sampling which we discussed in an earlier chapter apply in the selection of persons to be interviewed. Unplanned selection of informants, which is not to be confused with *random* selection in the statistical sense, usually does not yield a representative sample. Much interviewing by journalists is of this type. The readily available and the garrulous person gets an undue degree of attention in such interviewing. Indeed, the validity of results secured through interviews, as was also noted in connection with our discussion of mailed questionnaires, depends on the representativeness of the persons who cooperate. We saw in Chapter VII that the people who do not answer questionnaires frequently differ in important ways from those who do.[7] The interview, as contrasted with the mailed questionnaire, is sometimes assumed to avoid this difficulty. But obviously the interviewer will also fail to reach or secure the cooperation of some persons selected to constitute the sample. Do the people thus missed differ *in respects important for the particular investigation* from those

[6] For a discussion of *Principles Governing the Choice of Sources* in relation to social case work, see Mary E. Richmond, *Social Diagnosis,* (Russell Sage Foundation, 1917), pp. 169-76.

[7] A method of analysis for estimating the maximum effect of missing cases not found, unknown cases, and refusals is proposed by S. A. Stouffer and P. F. Lazarsfeld in *The Family and the Depression,* (Social Science Research Council, Bull. 39, New York, 1937), p. 176.

who are successfully interviewed? A recent study of this subject revealed the results shown in Table VII:[8]

TABLE VII

Some Personal Characteristics on which Refusals, Cooperators, and Unavailable Groups for an Interview Differed

Personal Characteristics	*Refused* %	*Cooperated* %	*Unavailable* %	*Total* %
Female	63.2	51.6	30.4	50.8
Low economic status	12.3	20.1	25.5	20.2
Low educational level	59.6	43.2	47.0	44.0
Industrial workers	21.0	28.1	44.1	28.9
Housewives	59.7	44.9	28.4	44.3
20 to 24 years of age	5.2	9.9	15.8	10.1
25 to 44 years of age	50.9	43.2	37.6	43.1
Number of cases	57	1635	102	1794

This study indicates that "refusals tended to come most frequently from housewives and people of low educational status. The unavailable group was predominantly male, included more industrial workers than other groups, and tended toward low economic levels. Refusals were characterized by individuals of intermediate ages, while more young individuals were unavailable for interviewing."[9]

2. *Securing appointment for interview.*—When the persons to be interviewed have been selected, the next step is to secure appointments with these persons. Securing a definite appointment is desirable from several points of view; it saves the time of the field worker by insuring him that he will be received, and by enabling him to use his time most economically; it avoids the antagonistic attitude which is frequently, and justly, aroused by the attempt of anyone to "break in" when the prospective informant is engaged in other work. Ordinarily, a request by letter for an interview

[8] Hazel Gaudet and E. C. Wilson, "Who Escapes the Personal Investigator?" *Jour. Applied Psychol.*, 24:775, 1940. This study is based on attempts to re-interview a representative sample of 1800 (out of an original sample of approximately 2800) by the Office of Radio Research, Columbia University. The panel was interviewed approximately once a month for the six months before the election of 1940. See Elmo Roper, "Checks to Increase Polling Accuracy," *Public Opin. Quart.*, 5:87-90, 1941.

[9] Gaudet and Wilson, *op. cit.*, p. 774.

is preferable to a request by telephone, as the former permits a more adequate statement of the purpose for which the interview is desired, the auspices under which the inquiry is undertaken, and other matters designed to insure the granting of the interview and a favorable reception. There are, of course, circumstances under which so complete a statement of the purpose of the interview would insure its *not* being granted, or cause a "preparation" on the part of the person to be interviewed which might materially destroy the value of the information secured. The technique of securing an appointment must depend on the nature of the information sought, the personal idiosyncrasies of the person to be interviewed, and the degree of rapport between him and the interviewer, or between the interviewer and the auspices under which the interview is sought. Sometimes interviews may be secured through the influence of friends or associates of the person with whom the interview is sought. Some form of preliminary arrangement for the interview is desirable when convenient, but much successful marketing and other research, such as public opinion polls, is conducted without special arrangement for interviews.

3. *Favorable conditions for interviewing.*—For most interviews privacy is highly desirable, though the importance of this varies with the nature of the interview.[10] When personal data are sought, the presence of other people is almost always inhibitory. For such interviews, the office of the interviewer acts as an inhibitor through the suggestion of officialdom. Where personal revelations are desired, places most removed from the social milieu of the informant yield the best results.[11] There are other types of interviews which can best be carried out in the office of the person interviewed, where records and corroborative material are available. The place and environment in which the interview takes place must be governed by the nature of the interview and the practical considerations involved.

[10] *Cf.* E. S. Bogardus, "The Social Research Interview," *Jour. Applied Sociol.*, 10:73, 1925. For certain purposes the group interview may be the most satisfactory and economical of the field worker's time. See Bogardus, "The Group Interview," *Jour. Applied Sociol.*, 10:372-82, 1926.

[11] *Cf.* Read Bain, "The Impersonal Confession and Social Research," *Jour. Applied Sociol.*, 9:359, 1925.

4. *Technique of the interview.*—When the interview has been secured, its usefulness will depend largely on the personality and skill of the interviewer. All the subtle influences of personal suggestion now become operative. If the purpose of the interview is merely to get certain objective facts which the informant has little or no reason to hide, no subtlety is required. The technique to be employed in a given case will also be determined by the general education, specific training, and personality traits of the informant. It is well for the interviewer to inform himself as far as possible on these matters before the interview and to prepare himself accordingly. But no amount of preliminary preparation will serve as a substitute for tact and spontaneous insight in a concrete situation, regardless of the type of interview. In the more difficult "confessional," "life history," or other self-revealing interviews these qualities are highly essential.

Thorough familiarity on the part of the interviewer with the purpose of the interview may be set down as a general requirement for a successful interview of any kind. Only through such familiarity can we know whether the informants actually get the meaning of the questions, and how we are to interpret their answers. Everything said in a previous chapter regarding the formulation of questionnaires also applies to interviews, although the latter afford additional opportunities of reformulating questions until they convey the desired meaning to the informant. This should not be taken to mean that less careful preliminary preparation is necessary in interviewing. The best phrasing of questions and the anticipation of possible misunderstandings should be carefully considered before going into the field.

The type of interview and the kind of information sought will determine whether a formal schedule can be used in the interview. Whenever possible, it is desirable for the interviewer to memorize both questions and answers so that writing can be dispensed with during the interview. The writing down of answers in the informant's presence may operate as an inhibitory influence in certain types of interviews. On the other hand, reliance on memory is subject to the possibility of inaccurate recollection of both questions and an-

swers. This is a problem which must be decided in each investigation on the basis of the type of information sought in relation to the quality and experience of the interviewers available. One of the subtlest sources of error which the interviewer must guard against is his own bias. Rice reports, for example, that an investigator who was personally a prohibitionist attributed 62 percent of the applications to a municipal lodging house to drink and only 7 percent to "industrial conditions." A socialistically inclined investigator found for a presumably similar random sample of applicants, 39 percent attributable to drink and 22 percent to industrial conditions. It is true that while the investigators are reported as experienced members of the staffs of social agencies, the questions of the schedule used were admittedly defective in allowing or possibly inviting the bias found. Other investigators under much more carefully controlled conditions have found, however, similar subtle influences of the interviewer's bias.[12] We have mentioned in a previous connection the dangers of leading questions, the tendency for people to give the kind of answer the interviewer wants, and the necessity for extreme care in the formulation of questions. These conditions must be even more carefully observed in the interview, because all the subtle influences of personal contact—appearance, facial expression, tone of voice, etc.—are here operative in addition to the more obvious influences.

The first step in the interview is to gain the confidence of the informant and to release whatever mental inhibitions may exist with reference to the interview. There are a great variety of devices in common use for this purpose. Ordinarily the interviewer will first introduce himself and refer to the local or other backing of the investigation. The first reaction of the interviewee is likely to be, "Here is another salesman." Accordingly, it is well to emphasize (*a*) that *information* only is sought; (*b*) that the informant's help is sought as an act of *generosity* on his part, rather than as something from which he or humanity will greatly profit—

[12] See Paul F. Lazarsfeld, "Repeated Interviews as a Tool for Studying Changes in Opinion and Their Causes," *Bulletin of the American Statistical Association*, 2:3-7, 1941.

the advancement of science, etc.; (*c*) that the information will be treated as confidential (when such is the case), and will never be used to the informant's disadvantage or embarrassment. It is well also to start with easy, conventional questions which no one objects to, or with questions of special interest or challenge to the informant.[13] To get him to talk at all, on any subject, is the first objective. Once he has started answering questions, he will answer many inquiries which would have been hopeless before he was "warmed up." Evidences of personal interest in the informant, aside from the purpose of the interview, may be helpful. The interviewer who has membership-character in the group or the class of people he is interviewing has obvious advantages in putting the informant at ease. As Katz has said: "A poorly dressed manual laborer with ungrammatical speech will have difficulty in interviewing along Park Avenue. Similarly, the Harvard graduate with a Harvard accent may not win the confidence of the miners in the Pennsylvania coal fields."[14] To evidence interest in the informant's occupation; to show one's own familiarity with his work; to refer to important friends of the informant as if one were quite well acquainted with them; to tell of one's own experiences or problems and ask the informant's advice or reaction to them—these are illustrations of the scores of devices employed to get an informant "started."

Indirect approaches may sometimes yield more valuable results than direct requests. Likewise, impersonal and disinterested demeanor will do more to secure frankness than will direct urging to be frank. When there is reluctance to communicate important facts, the interviewer may stimulate communication by appearing to be misled in a direction disadvantageous to the informant. This will sometimes bring forth the desired information in the form of a correction. The extent to which these devices are useful is, again, determined by the type of informant. There are objectively minded people who are glad to give information for the sake of truth and the advancement of research, and who would

[13] See Chap. VII, pp. 189-90.

[14] D. Katz, "The Public Opinion Polls and the 1940 Election," *Public Opin. Quart.*, 5:52-78, 1941, p. 69.

resent any but the frankest and most direct approach. The technique of the interview must be modified to suit each occasion.[15]

Bogardus summarized the methods of interviewing and the types of informants as follows:

Experience indicates that a witness does best when (1) allowed to use a narrative form, (2) when least interrupted, (3) when encouraged to follow a time order, and (4) when answers are not suggested by questions, but when questions are used simply to release narratives. Persons display a variety of ways in "witnessing," and hence the examiner must change his methods accordingly. (1) The rambling witness requires a close and catechetical examination. (2) The dull and stupid witness calls for inexhaustible patience on the part of the advocate. (3) The timid and self-conscious witness needs to have his attention taken off the effects regarding himself or others that he imagines are going to be produced. (4) The bold and zealous witness is easily led astray by his own exaggerated egoism and is usually a dangerous witness from the standpoint of truth. (5) The hostile witness is of value to truth only in indirect ways, by studying his negative reactions — otherwise, he is to be avoided. (6) The lying witness may be (a) an innocent liar, (b) a careless liar, or (c) a wilful liar. The classification suggests in itself the attitude to take toward him. (7) The flippant witness is to be circumvented, to be treated as a "wild animal ready to tear you if she should get near enough." Her "frenzied exuberance" is to be encouraged, for thereby she will "give herself away." (8) The dogged witness is to be approached by "getting little answers for little questions," and to get him loosened up, and accustomed to talking. (9) The hesitating witness is to be allowed to take his time. His weighing and balancing of answers needs to be studied. Sometimes he may hesitate simply in order to be "scrupulously accurate." (10) The nervous witness is to be dealt with gently and to be encouraged. (11) The humorous witness may have his good humor appealed to, and may be as he is, a jolly good fellow. (12) The cunning witness is not cunning so much as he thinks he is. He is to be met with cunning and needs to have his real character "shown up." (13) The canting hypocrite is likely to believe in "religion and his own goodness," and to approach a down-right lie by shirking it and by using phrases that disclose his weaknesses. (14) The positive witness can be led into contradiction of what she or he has already said.[16]

[15] *Cf.* E. S. Bogardus, "The Social Research Interview," *Jour. Applied Sociol.*, 10:69-82, 1925. See also E. T. Krueger, "The Technique of Securing Life History Documents," *Jour. Applied Sociol.*, 9:290-98, 1925.

[16] E. S. Bogardus, "Methods of Interviewing," *Jour. Applied Sociol.*, 9: 459-61, 1925.

As stated at the outset of this section, it is not possible here to enter into a detailed consideration of the intricate interstimulation and response which are the structure and content of the interview. The fact is that there are very few scientific data available on the subject, although research in this field lies at the very foundation of sociology.[17] Like all other techniques, skilled interviewing and social work generally must wait upon the development of the basic sciences which underlie these techniques. Only from such foundations can a scientific technique be developed.

Skillful interviewing will remain an art. As in the case of other arts it will become a more refined and communicable technique in proportion as the scientific principles underlying it are understood and applied. These principles must in turn evolve largely from the experience of successful interviewers. Pauline Young has summarized that experience, especially as it relates to social case work, in the following practical maxims:[18]

I. Preparation for the Interview:
   1. Understand why the interview is held.
   2. Be prepared to be viewed by the interviewee and to answer any question he may wish to ask.
   3. Gather all possible clues and concentrate on each one to see which might be best utilized toward a successful approach to the interviewee.
   4. If time is limited and problems are pressing, formulate tentative objectives for your interview or chart its course as far as possible.
   5. Keep in mind the unique problems of each case and relate them to the total social situation in which the interviewee finds himself, as far as is possible under the circumstances.
   6. "Size up" your interviewee from whatever clues you may possess, but keep your mind flexible when you are in contact with the interviewee.

[17] The basic behavior involved has been approached in a detailed quantitative way by E. D. Chapple, "Measuring Human Relations: An Introduction to the Study of the Interaction of Individuals," *Genetic Psychology Monographs*, 22:3-147, 1940. For a general bibliography of the subject, see Pauline Young, *Interviewing in Social Work*, (McGraw-Hill, 1936). For a keen analysis of some psychological aspects of interviewing, especially as related to market research, see *The Technique of Marketing Research*, (McGraw-Hill, 1937), Chaps. 3 and 4, by Paul F. Lazarsfeld.

[18] Pauline Young, *op. cit.*, pp. 86-88.

7. Make appointments whenever possible, thus showing respect for the client's time.
8. Provide privacy, comfort, and agreeable surroundings.
9. Be modestly and suitably dressed for each occasion.
10. Introduce yourself to the interviewee, explain the function of the agency you represent, and state briefly the purpose of visit.
11. Extend cordial greetings.
12. Observe conventionalities of the interviewee's home and social group.

II. The Interview Proper:

13. Show interest in everything the interviewee wishes to tell you. Make the interview "his moment."
14. Be at ease with the interviewee, thus helping him to be at ease with you.
15. Adopt a leisurely pace to give the interviewee the feeling that you consider his situation seriously.
16. Allow sufficient time to learn the facts which will enable you to understand the problems involved.
17. View all problems confronting the interviewee from his point of view.
18. Make suitable allowances for your preconceived notions about his problems, for your biases and prejudices.
19. Meet the interviewee on his own terms.
20. Gain, keep, and deserve the interviewee's confidence.
21. Identify yourself with the interviewee through similar experiences, points of view, mutual likes and dislikes.
22. Gain sympathetic insight into problems confronting the interviewee.
23. Assign social status to the interviewee, remembering that status is one of his most priceless possessions.
24. Ask only questions which you would not resent answering about yourself under similar circumstances.
25. Ask questions which are easily understood.
26. Ask questions which do not imply their answers.
27. Take care not to sound insinuating or impertinent.
28. In asking questions be frank and straightforward rather than cunning.
29. Avoid "getting around" the interviewee or extracting information against his will or knowledge.
30. Allow the interviewee to tell his story in his own way and to take his time in telling it.
31. Allow the interviewee to think for himself and give him time to think his situations through.
32. "Make your minds meet"; that is, be sure that you understand what the interviewee says, what he wants, and what his culture complexes are, and he yours.

33. Listen with interest.
34. Follow up every important clue of interest to the interviewee.
35. Seek to understand source, cultural setting, and development of the interviewee's behavior patterns.
36. Meet objections in a way which is satisfying to the interviewee.
37. Avoid ordering-and-forbidding techniques.
38. Allow for face-saving.
39. Lessen tensions by assigning status, by soliciting help, by complimenting good performance, etc.

III. Controlling the Interview:

40. Check the story in your own mind to see if there are any inconsistencies.
41. Deal with inconsistencies as misunderstandings.
42. Remember the saying that truth is told to friends and lies to enemies.
43. Eliminate opportunities which an interviewee might seize upon to deceive you.
44. If you sense that an interviewee is going to falsify, start out by stating the facts of the case as far as they are known to you.

IV. Closing the Interview:

45. Close the interview before you have worn out your welcome.
46. Close the interview when the interviewee is at ease emotionally.
47. Close the interview when the interviewee has caught the desire to manage his own affairs.
48. If subsequent interviews are necessary, close when you still have something important to talk about.
49. At the close of the interview ask if there is anything else your interviewee wishes to talk about. "What else would you care to tell me? What didn't we touch upon?"
50. Test the success of your interviews.

We have discussed above some principles of interviewing which may be said to have general validity, but which are primarily adapted to research rather than therapeutic or other special purposes as regards the informant. Specialized forms of the interview such as concern legal, medical, psychoanalytic, and general social work purposes are beyond the scope of the present discussion. Two special types of

interviewing, namely, the group interview and the repeated interviewing of a "panel" should be mentioned in passing.

E. THE GROUP INTERVIEW

For some purposes an oral interview with a group of people may yield information of a corroborative or "consensus" type which is perhaps more reliable than individual interviews with each member would have been. The tendency of individuals to dramatize and to speak positively of matters about which they know little is likely to be restrained in the presence of other people who know about the subject. The people interviewed in a group may call each others' attention to aspects of a situation which receives general assent when mentioned, but which probably would not have occurred to more than one or two people in individual interviews. Obviously, the group interview is suited to securing only certain types of information. When a suitable subject is under investigation, however, and the right kind of group can be assembled, the method has obvious advantages in taking less time than individual interviewing.

F. THE "PANEL"

The labor and expense involved in large scale interviewing, and especially in the selection of a new sample for each inquiry, has resulted in the so-called panel technique. This method consists of carefully selecting a representative sample of population and securing their cooperation for repeated interviews, as a sort of permanent body of informants. The technique is especially useful when a great wealth of detailed material is sought with special reference to changes over a period of time. Thus, a nation-wide sociological study called Mass Observation was begun in England in 1937.[19] Only the briefest summary of the procedure is possible here:

On February 12, 1937, thirty people who did not know each other, who were in widely-scattered areas, and who were of widely-ranged ages and income levels, wrote down everything that occurred to them on that day in as great detail as they could muster. At about the same time, six observers went to a Lancashire cotton

[19] See *First Year's Work*, by Mass Observation, edited by Charles Madge and Tom Harrison, with an essay on "A Nationwide Intelligence Service," by Bronislaw Malinowski (Univ. of London, Lindsay-Drummond, London, March, 1938).

town where they devoted all of their time to participation in and description of the lives about them. Another group went to another town where the members of the former community were accustomed to spend their holidays. The first group of thirty, plus hundreds more who later volunteered to join their ranks, recorded impressions, opinions, and actions; in short, everything they noticed in themselves or those around them. The two groups that entered into the life of the community were instructed to observe in much the same fashion but were also given instructions from time to time as well as elaborate questionnaires, dealing with specific factors of smoking, drinking, gambling, and the like. There were, then, three kinds of reports: day surveys on the twelfth of each month made by observers all over the country, area surveys consisting of detailed reports on community life by full-time observers, and studies of isolated factors.[19a]

Strictly speaking, the first group is the most analogous to the panel, since the day surveys consist of periodic reports by the same group of people concerning their own opinions and activities. The full-time observers within given communities are more comparable to trained interviewers or to what the anthropologist calls "informants." If the community is so small that each member of it is interviewed about various aspects of his life, then the community itself forms another panel. If surveys of specific factors such as smoking and reading habits, on the other hand, are given to a random sample of the general population as well as to the day-survey people and the community observers, it is hoped that these two types of material, the unselective detailed mass of day surveys, and the controlled interview-questionnaire may eventually be used to check as well as to supplement each other. Such a plan depends for its success on the sifting and classifying of the day-survey material by trained research workers and clearly-determined points of reference. The workers are becoming more and more skilled, but there is increasing need for a structural index of the material which will be the basis for transforming a mass of subjective material into objective data.[20]

The panel technique has also been used by such organizations as the National Board of Motion Picture Review, *The Woman's Home Companion*[21] and by the Office of Radio Research.[22] The director of the latter organization has sum-

[19a] There is no mention of how these questionnaire respondents were selected or interviewed; they may have been observers, members of the two communities studied, a random sample, or all three. For further critical comments on the method, see R. Firth, "An Anthropologist's View of Mass Observation," *Sociol. Rev.*, (London), 31:166-93, 1939.

[20] Paul Lazarsfeld and Marjorie Fiske, "The 'Panel' as a New Tool for Measuring Opinion," *Public Opin. Quart.*, 2:596-612, 1938.

[21] *Ibid.*, pp. 5-10.

[22] P. F. Lazarsfeld, "Panel Studies," *Public Opin. Quart.*, 4:122-28, 1940.

marized admirably the advantages and problems of the panel technique as follows: [23]

1. If, in a special study, one is interested in more than a straight "yes" or "no" answer to a question, and needs elaborate comments by the respondents, it takes considerable time to get the cooperation of people and to improve their ability to express themselves. If such an effort has been successful with one group, one can capitalize on it in each subsequent interview. For example, it is hardly possible to use a different group of people every week if detailed comments on a radio program or on issues of a magazine are needed.

2. If one is interested in relating the opinions of people to their *personal characteristics,* the repeated interview gives us an opportunity to accumulate knowledge regarding the correspondents. It would be hardly possible in one interview to ask people about their leisure-time activities, give them a personality test, collect biographical information, and so on. But it is possible to secure part of this information each time we return to the panel.

3. Under certain circumstances, the *statistical reliability* of repeated interviews with panel members is greater than that of answers gained from a series of distinct samples.[24] A change of a few percent might be reliable in the former, but unreliable in the latter case. The panel, therefore, can be smaller than the serial sample. It is also technically easier to get good representation in one panel than to get the same effect with a great number of samples.

4. In studying the attitudes of people toward an event which is itself *extended in time,* the panel is probably the only adequate tool of investigation. . .

5. In certain cases, the panel is a necessary *substitute for the use of a control group,* which latter is not as dependable a tool of research as is generally assumed. The use of control groups is taken from the physical sciences. The student of agriculture, for example, applies fertilizer to one part of a field and does not apply it to the other part of the field. The difference in yield between the two parts, if reliable, can be safely taken as the effect of the fertilizer. In social research, however, the situation is not so simple. If we compare, for example, employed and unemployed people as to their political attitude in order to see what the political effects of unemployment are, we cannot assume that the employed people are an adequate control group for the unemployed. The control would be adequate only if we could take a number of people, throw half of them out of their jobs, and then see how their political attitudes change, compared with

[23] Paul Lazarsfeld and Marjorie Fiske, *op. cit.,* pp. 1-5.

[24] For a demonstration of this point see P. F. Lazarsfeld, "Panel Studies," *op. cit.*

the employed group. But in any concrete research situation, the "control group" might have become unemployed for reasons which themselves affect the political attitudes. Most of the control groups available for social research are "self-selected" in this sense. By comparing, for instance, farmers who have radios with farmers who do not have radios in a rural non-electrified area, the latter group would prove to be, at the same time, the poorer farmers who cannot afford batteries. Therefore, any comparison between the two groups might pertain to economic differences much more than to the influence of radio ownership.

Under certain conditions the panel is preferable to the control group method. . . It might have been possible, for instance, to ask a group of people how they felt about the Reorganization Bill before a speech by the President, and then repeat the question immediately after the speech had been delivered. Certainly this procedure would be more valid than comparing the opinions of people who have and who have not heard the President because in the latter case the opinions might influence the willingness or unwillingness to listen more than the listening might influence the opinions.

6. The *saving in cost,* which the use of a panel would mean, is quite obvious and has already been mentioned.

The problems involved in using the panel are closely related to the advantages. Once people have been induced to cooperate in a panel, their patience can be taxed much further. But how difficult is it to get people's participation? How great is the mortality even among those who have participated?

It is further probable that people who join a panel grow more articulate than those whom we approach by mere doorbell ringing for a short poll questionnaire. But not all of them will be articulate and if we use the opinions of those who talk much, don't we introduce a new fallacy? The problem of articulateness is not peculiar to the panel, but it does not enter into a poll where we ask only very simple questions. It becomes an essential problem in a panel, however, when it is set up for just this purpose. . .

A poll sample requires only that it be representative for the whole population of interest and large enough for the returns to be reliable. A jury introduces a new element. It has to be representative too, but it differs from the poll sample by the fact that the individual approaches the test knowing that he will be called upon for an opinion. To give a concrete example: in a poll we ask people how they liked the speech they heard yesterday—in a jury we ask the members to pass judgment on a speech they will presently be hearing. This "critical set" is a problem for the panel method as well as for the jury method. The panel also has to be statistically representative and large enough, just as poll and jury. The additional problem introduced by the panel

is that we do not know what becomes of this "critical set" as a result of the repeated interviews. Do people become more and more critical—drift, so to speak, into a professional attitude—more and more divorced from the life situation of the general public? Or must we reckon with an increasing "freezing" of the attitudes of the panel members? This freezing could conceivably occur with, for example, people who have once explicitly stated their political attitude and are consequently less likely to change their minds than the general public. It might be, also, that people who have repeatedly tasted a manufacturer's product become less discriminating.

So it appears that the use of the panel as a measurement of public opinion holds great promise; but it also opens new areas for research which must be made before it can be used with safety.

G. THE PARTICIPANT OBSERVER

We have been concerned above with the more or less formal interview in which the recognized interviewer-interviewee relationship exists. Some of the most enlightening field work has been carried out by means of a far more intimate and informal set of relationships between the investigator and the investigated, namely, through the so-called "participant observer" technique. It consists of the observer's becoming as nearly as may be a member of the group he investigates. That is, he settles in the community and participates in the everyday life of the group. The degree to which this is possible will vary with the characteristics of the investigator in relation to the culture of the community he has chosen to investigate. There are the instances of explorers like Knud Rasmussen[25] and Peter Freuchen[26] who settled down among the Eskimos and adopted their ways of life for periods of several decades. Freuchen, for example, married a native woman, reared children, and in all other ways shared the Eskimos' life. It is difficult, of course, to distinguish between fact and fiction in some travelogues and other accounts of adventurous journalists and lecturers. But when men of scientific training and ideals produce such accounts, they may reveal aspects of a culture that more formal methods can not portray.

Less extreme forms of the participant observer technique

[25] Knud Rasmussen, *Across Arctic America*, (Putnam, 1927).
[26] Peter Freuchen, *Arctic Adventure*, (Farrar and Rinehart, 1935).

combined with more formal inquiry are such investigations as the Lynds' *Middletown,*[27] in which the investigators participate for a period of some weeks or months in a culture very much like their own. In any case, the principle is to observe by sharing as nearly as may be in the behavior observed. It is an attempt to reduce to a minimum the distortion of actual situations which may arise from the fact that the observer is an outside agent.[28]

A good critical account of the participant observer technique in social research is Florence Kluckhohn's summary of her experience in a Spanish-speaking village in New Mexico.[29] The student should study her paper in full. We can reproduce here only a few extracts.

> In the application of the technique it is essential to realize that it is not only the investigator himself but also the members of the group being studied who are to regard him as a participant. This means that he must achieve status within the community organization, must play his part in a constellation of roles. . . The basic factors which determine general roles are such objective states or conditions as sex, age, position in a caste or class, and race. . .
>
> The would-be participant must learn something of the manner in which these determinants of general roles are dealt with in the social organization of the particular community in question. If a general knowledge of the social organization is not obtainable from studies of similar communities, there must be a period of watchful waiting. . .
>
> The investigator is never able to shake off entirely his role of outsider, and I am in accord with those who maintain that it is not advisable for him to do so. Some exceedingly valuable information comes to the outsider simply because he is one. It is the investigator's task, however, to fashion all his roles, including this one, into tools; and the role of outsider can be an excellent device if one is in command of it. The critical point here would seem to be whether the community categorizes one as a particular kind of outsider or just vaguely an outsider. It is once again a matter of orientation. . .
>
> What is the value of being a participant-observer? First, it

[27] R. S. and H. M. Lynd, *Middletown,* (Macmillan, 1929), p. 506.

[28] The writings of C. H. Cooley on "sympathetic sharing" of experience in social research have never been surpassed for clarity of exposition of this subject. See especially *Human Nature and the Social Order,* (Scribners, 1902), Chap. 4. Also *Sociological Theory and Social Research,* (Holt, 1930), Chaps. 9 and 10.

[29] "The Participant Observer Technique in Small Communities," *Amer. Jour. Sociol.,* 46:331-43, 1940.

would seem probable that the opportunities for observation are increased and that in some respects greater objectivity and accuracy insured. Second, it would appear that a use of the technique will provide a desirable balance between the more purely behavioristic type of investigation and the type which seeks some measure of insight into the "meanings" current in the community.

In treating the technique comparatively its advantages and disadvantages are considered in terms of what we shall call the three "R's" of evaluation, the *range, relevance,* and *reliability* of the data obtained. Does the technique increase the field of selection? Does it provide data which are more relevant to the interests and activities of the group in question? . . . And does it provide more reliable data than certain other techniques?

In four rather obvious respects the range of data would seem to be increased by use of the technique. First, it affords access to the data which come from observation in the current situations in which the community members are involved. The possibility of a large and consistent body of such information is peculiar to the technique in question, for, if the investigator seeks to gain information by the direct interview, the questionnaire, or the life-history, he is creating special situations, not participating in those which arise within the group itself. Interviewing indirectly while helping a woman with her washing is very different from interviewing directly in the woman's parlor or one's own temporary office. Even though the answers to the questions may be the same, which is doubtful, what one observed is far from the same.

To some extent, of course, even the participant observer has an effect upon situations. He cannot wait, for example, for persons to recall voluntarily past events or to explain the relationship of A to B. He asks leading questions at some points and, in doing so, alters the situation, but he is altering given situations in varying degrees, not creating special situations.

The range of available data is likely to be increased in a second respect. There are in all groups certain kinds of data which are guarded more closely than others. Direct questions regarding such information may be met with evasions if not outright misrepresentations. Indirect questions may also fail. Simulation of behavior made possible by participation may, however, open the door to this guarded realm. Space does not permit illustration of this point, but to indicate what is meant: It was through a gradual simulation of the Spanish-American woman's fear that I obtained most of my information about witchcraft and other beliefs which are jealously guarded by the Spanish-Americans.

The ready access to gossip is the third respect in which participant observation serves to increase the range of information. Gossip is an important source of information both for what is said and for what is not said. The comments made about be-

havior often give checks to and judgments upon behavior which can be obtained in no other way. Moreover, it is especially important to listen to gossip if one wants to discover the patterns of "socially approved" behavior. A bit of behavior is observed. It may strike the outsider as unusual. If community members begin to comment upon it, and especially if they condemn it, it is relatively safe to assume that it is unusual behavior for the community in question. If, however, there is no comment whatsoever, it is also reasonably safe to assume that it has not struck community members as being deviant behavior.

The fourth respect in which range is increased has already been suggested. It lies in the advantage derived from being in the position of the person who is being taught.[29a] It has long been an argument of field research workers that the prestige of being an outsider is of great importance, that people talk because they are flattered by attention from an outsider. This does apply to a certain type of person, but it may be added that all people, even the simplest, have "egos." While some individuals are flattered by attention and are happy for the opportunity to talk at length, the egos of research workers themselves have led to over-emphasis of this factor and to too little regard for the detrimental effects of a suspected attitude of superiority and condescension. . .

There are, of course, some disadvantages in participant observation. Participation in a community's general roles may well restrict the range of observation and interview in one direction while increasing it in another. In a rigidly stratified community it may be disadvantageous inasmuch as participation in one segment may preclude participation in another. Even in relatively homogeneous communities there is always some restriction. Participant observation may also force one to sacrifice a certain type of uniformity in data. It will not be possible to put a given question to a large number of people — at least not in the same form. Attitude tests, for example, or house-to-house inquiries on various matters cannot be made. For those problems where such inquiries are essential participant observation will be a hindrance rather than a help. But in small communities such information can often be obtained indirectly and more accurately, for there will be less danger that surface uniformities are obscuring subtle differences.[29b]

In short, different problems require different methods of approach, and different approaches yield different types of

[29a] Cora Du Bois, "Some Psychological Objectives and Techniques in Ethnology," *Jour. Sociol Psychol.*, 3:285-301, 1937.

[29b] George A. Lundberg and Margaret Lawsing ("The Sociography of Some Community Relations," *Amer. Sociol. Rev.*, 3:318-36, 1937, especially pp. 326-27) point to certain limitations of the questionnaire method in procuring friendship patterns which would seem to bear out this point.

results. All of them are valuable for their respective purposes, and with respect to the type of result sought. Their combination yields the fullest description of the many-sided phenomenon of social behavior.

H. THE LIFE HISTORY OR CASE HISTORY

The modern life-history document is a lineal descendant of very ancient methods of sociological description and generalization, namely, the parable, the allegory, the vision, the revelation, the myth, the story, and the novel.[30] The modern life history, however, is a detailed voluntary account of a person's own view of the events of his life as revealed under the sympathetic questioning of a skilled inquirer. The objective facts of such a story may be corroborated by the usual historical methods. But the life history aims to reveal not only these *facts* but their *meanings* to the informant. It is these meanings that are important in explaining his behavior. The illuminating character of life-history records, both as reflections of the general character of different cultures throughout the course of history and as the source of hypotheses regarding human personality and social behavior, is recognized by everyone. As scientific material, much past and contemporary life-history material has been questioned on account of the informality and subjectivity of the observations. Yet such studies frequently describe the sequence of adaptive processes throughout the life of an individual, and in this way have great value in practical and therapeutic adjustments as well as in providing a type of sociological knowledge. These values make it quite out of the question to dispense with such records, whatever their scientific limitations may be.

Accordingly, recent students of the problem have properly devoted themselves to the development of techniques for the recording and analysis of life histories which will more adequately meet the scientific standards of observation and generalization. There is no reason why this should not be successfully accomplished. The assumption that there is any fundamental incompatibility between the life-history

[30] *Cf.* L. L. Bernard, "The Development of Methods in Sociology," *The Monist,* 38:292-320, 1928.

"method" and the techniques with which this book is mainly concerned is clearly unwarranted. "As a matter of fact," says Kimball Young, "many of the basic problems of [case history] method are not unlike those of statistics. There is, of course, a recognition of a set of specific or general factors to be studied. The more careful case-history studies also recognize the problem of sampling, since the aim is to avoid the temptation to generalize from one case. So, too, references are frequently made to typical cases, which is a way of referring to some central tendency. Likewise, attention is given to the divergent instance, which is but an informal way of stating facts about variability. Finally, the worker with these data may compare his findings with those from other samples, and in order to expose dynamic relations he may indicate in qualitative terms covariation or correlation between selected factors or situations." [31] Bernard has stressed the same point in the following passage:

> The real source of methodological error [in case study] is in supposing that statistical generalization can proceed without definition of data or units or that generalization can proceed from case data without some sort of statistical method. So-called commonsense inferences from cases amount to nothing less than informal statistical generalizations. Formal statistical generalizations are more adequate, when practicable.[32]

The problem of how to observe, classify, and generalize life-history material remains largely to be solved, but there is no reason to consider the problem insoluble. Important steps have already been taken toward the establishment of criteria for the more adequate recording of such material and the development of methods of generalizing from the material secured. Space does not permit us even to summarize the techniques of securing intimate personal life histories through protracted interviews of the type employed by psychoanalysts, psychiatrists, sociologists and the abler social case workers. A general idea of the more systematic approach may, however, be secured from John Dollard's standards for an adequate life-history technique:

[31] Kimball Young, *Personality and Problems of Adjustment,* (F. S. Crofts and Company, 1940), pp. 250-51.

[32] L. L. Bernard, *op. cit.,* pp. 306-7, 314.

I. The subject must be viewed as a specimen in a cultural series.

II. The organic motors of action ascribed must be socially relevant.

III. The peculiar role of the family group in transmitting the culture must be recognized.

IV. The specific method of elaboration of organic materials into social behavior must be shown.

V. The continuous related character of experience from childhood through adulthood must be stressed.

VI. The "social situation" must be carefully and continuously specified as a factor.

VII. The life-history material must itself be organized and conceptualized.[33]

The full implications and applications of these standards in actual cases have been discussed and illustrated in detail by Dollard. As in all research, the best grasp of a method is to be secured from a study of its actual use in conspicuously successful research. (See references at the end of this chapter.) Space permits us here to touch only on the aspects of the life-history method which have special relevance to interviewing.[34]

What has been said in a previous section about the importance of a sympathetic relationship between the interviewer and his informant, is, of course, of special importance in the life-history interview. The images which both participants in the interview have of each other, and the different roles which each assume will fundamentally condition the subtle interaction of personalities which characterizes especially this type of interview. It is this interaction which the interviewer must be able to observe and report objectively, although he is himself an integral part of the situation. Kimball Young has summarized the matter as follows: [34a]

In identifying himself with the informant the interviewer must be able to sense or "feel" the former's aims, traits, attitudes, values, and other inner states. Yet in this understanding of the

[33] John Dollard, *Criteria for the Life History*, (Yale Univ. Press, 1935), p. 8. For some additional suggestions, see K. Young, *op. cit.*, pp. 250-58.

[34] For an outline of the different types of life-history materials, see Chap. VIII, pp. 220-21. The analysis of life-history documents is, of course, subject to all the principles of criticism treated in our summary of the historical method (Chap. IV).

[34a] K. Young, *op. cit.*, p. 257.

role of the other, the interviewer must not overdo his identification and thus lose his own integrity and his critical capacities. The successful interviewer must have the ability to play a dual role—his own as manipulator of the situation and seeker of information, and also in imagination that of the subject or informant. Only in this way will he be able to follow the latter's answers, comments and confessions sympathetically. In other words, he must learn to play another's role sympathetically in imagination, but he must not completely absorb such a role in his overt or verbal reactions. . .

The competent interviewer must possess keen perceptive faculties and an accurate memory. Not only should he hear correctly what is said to him, but he will be alert to the overtones of the informant's verbal and overt reactions, noting changes in voice, indications of feeling-emotional states from facial or other gestures, and any other possible clues to inner states which may subsequently be exposed. And, while he must know how to direct the session with skillful and revealing inquiries, he must also possess the capacity for sympathetic listening. Often in interviews, especially with persons under the emotional stress of mental conflicts, once the informant has begun to pour forth his troubles, the interviewer need but listen attentively in order to gain significant information about the fundamental inner life of his subject.[35]

In summary, we may say that the life history in some form or other remains the great common-sense source of information about social behavior. Nearly all of the social knowledge by which the average person makes his adjustments, as well as the overwhelming proportion of the knowledge on the basis of which social agencies carry on their work, consists of informally generalized case histories. The principles thus derived, with all their shortcomings, owing to inaccurate observation, inadequate sampling, and faulty inference, are unquestionably the guiding tenets by which most men live and much social administration is carried on. Scientifically derived generalizations have not as yet supplanted to any great degree the accepted conclusions from novels and voluminous but unsystematically generalized case histories. Representing, as these histories do, a kind of record of experience, it is reasonable that in the absence of more reliable information and principles, the well-known *case*, although it may be chiefly a stereotype, should be highly valued. At

[35] On the art of listening, see N. Diehl and R. S. Wilson, "Can Listening Become a Case Work Art?" *The Family*, 14:99-105, 1933.

the same time, the fallacy of using a mere illustration as proof is a very common and mischievous one.

In the meantime, important work is going forward in the production of more systematic, reliable, and relevant case records. We have noted the great value of life-history material in suggesting hypotheses for more rigorous scientific tests, in throwing light on past situations giving rise to present conduct, and in a great variety of practical adjustment situations. We have also noted the tendency of case studies to neglect the known principles of valid sampling, generalization from inadequate data, and the lack of standardized and objectified procedures permitting corroboration of both data and generalizations. We also have to consider the possibilities of errors in perception and reporting, failures of memory, unconscious omissions or repression of the unpleasant, distortion of recall, the tendency to dramatize and describe a desired imaginative role, and the temptation to project upon his subject the interviewer's own preconceptions, and to force the data into the theory with which we approach our task.

Many of these dangers are common to all scientific work, and cannot be considered peculiar to the life-history approach, although here some of them are undoubtedly especially difficult to avoid. Science must proceed, however, on the hypothesis that none of these difficulties are insuperable. The improvements already noticeable in the life-history techniques are gratifying and promising. As we proceed with this work the apparent differences between the life-history technique and other methods will tend to disappear. As Robert E. Park said many years ago regarding generalization from case histories: "The procedure here is just the same as in any of the natural and physical sciences. We explain them by putting them under some general category, classifying them, in short, and then discovering where we can, the reason for the deviation." [36]

There is no doubt that the above is a correct characterization of what is actually done in the case of life histories, as well as in the case of records of other behavior segments how-

[36] R. E. Park, "Experience and Race Relations," *Jour. Applied Sociol.*, 9: 22, 23, 1924.

ever secured and of whatever degree of complexity. The frequently assumed fundamental difference between case history and other sociological methods is, therefore, undoubtedly exaggerated. The different methods sometimes achieve the same result.[37] The main difference lies in the informality, and therefore in the verifiability, of the steps by which generalization from life histories is carried out. Since a clear account of the processes by which we reach our conclusions is a principal requirement of scientific method, this remains the chief problem in the use of life histories as scientific material.

The logic of generalization from case studies has been trenchantly considered by Stouffer and Lazarsfeld.[38] They find that "*rigorous* treatment of case history materials implies dependence on statistics and tends to be almost futile if a multiplicity of variables is involved." To reduce the numbers of variables, the device of typological classification is suggested. Such classification in its most elaborate mathematical form takes the form of multiple factor analysis as developed by Thurstone, Hotelling, Kelly, and others. The discovery of a high correlation between two or more variables makes it practically possible to neglect certain combinations of attributes or to reduce them to clusters which may then be treated as single variables. Thus Chapin in the *Social Status Scale,* 1933, found that a large variety of attributes having to do with socio-economic status could be reduced to seventeen items. Burgess and Cottrell applied, with suggestive results, such methods to their elaborate case records of different degrees of marital adjustment. The more refined methods of factor analysis are too involved to permit exposition here, but their possible contribution to the solution of the most difficult problem in the scientific utilization of case histories should be noted. The tendency in case re-

[37] S. A. Stouffer, *Experimental Comparison of Statistical and Case History Methods in Attitude Research,* (Ph. D. thesis, University of Chicago, 1930). For a brief summary of this study, see the *Publ. Amer. Sociol. Soc.,* 25:154-56, 1930. See also Ruth S. Cavan, P. M. Hauser and S. A. Stouffer, "Note on Statistical Treatment of Life History Material," *Soc. Forces,* 9:200-03, 1930.

[38] S. A. Stouffer and P. F. Lazarsfeld, *Research Memorandum on the Family in the Depression,* (Social Science Research Council, Bull. No. 29, 1937), Appendix A.

cording is to reduce the data to the form of an elaborate schedule or questionnaire.[39] The data thus lend themselves to statistical treatment. The argument that such procedure destroys the important "configurational" aspects of the data will probably be met in the end by typological methods involving the fundamental logic of factor analysis.

Finally, there should be mentioned the practical suggestions of Lazarsfeld and Robinson[40] for the more precise and quantitative classification of case studies. The proposed procedure seeks to formalize and render more precise the process by which discursive, literary case stories are now informally classified and evaluated. Consider, for example, the selection of an employee on the basis of a great deal of information about him. "One observes that the applicant speaks positively and seems to know his business but also that he avoids making decisions, and one estimates his desirability in the light of these conflicting indications. The selection of another applicant, however, may involve very different indicators; for example, the way in which a man expresses himself in a letter and the recommendations he encloses."[41]

The method suggested by the above authors is as follows:

A. *Define the continuum in terms of which the case studies are to be classified.* This requires an explicit statement as to the general quality or characteristic according to which the individuals will be classified.

B. *Decide which indicators* (i.e., "bits of information") *in each case study are items on this continuum,* i.e., *are relevant to the classification.* If the continuum is that of dominance-submission, any specific detail helpful in deciding whether an individual is dominant or submissive is an indicator on the continuum for that particular individual. Deciding which indicators are relevant to the classification thus consists merely in specifying or concretizing the definition of the continuum in terms of specific items of information.

C. *Give a numerical value, with sign, to each indicator, according to its position on the continuum.* To continue the preceding illustration, if a particular detail indicates that the in-

[39] E. W. Burgess and L. S. Cottrell, Jr., *Predicting Success or Failure in Marriage,* (Prentice-Hall, 1939), Appendices B and C. See also Chap. 16 of this excellent monograph for a discussion of the problem here under consideration.

[40] P. F. Lazarsfeld and W. S. Robinson, "The Quantification of Case Studies," *Jour. Applied Psychol.,* 24:817-25, 1940.

[41] *Ibid.,* p. 819.

dividual is submissive, it might be designated by — 1, to show that so far as this particular detail is concerned the individual is on the negative side of the dominance-submission continuum. If the specific detail indicates that the individual is dominant, this fact might be indicated by a + 1. A detail relevant to the classification, but indicative that the individual is neither very dominant nor very submissive, might be given a value of zero. A more detailed rating scheme might also be used if desired.

D. *Combine the "scores" on the different indicators for each case study in order to determine a final index of the positions of the case on the continuum.* The most efficient combination of the indicator scores for a single case is their arithmetic mean. If different numbers of indicators are found in different case studies, these means will of course be based on different numbers of observations. Thus if in a particular case study three positive, one zero, and four negative indicators are recorded, the final index will be $\frac{(+3)+(0)+(-4)}{8} = -\frac{1}{8} = -.125$. If for another case study there are four positive, two zero, and six negative indicators, the final index will be $\frac{(+4)+(0)+(-6)}{12} = -2/12 = -.167$. Computing indexes in this way allows the utilization of all relevant information found in the case study; none need be thrown away. The total range of the scores will, of course, lie between — 1 and + 1.

In combining the scores it is neither necessary nor desirable that the number of indicators be the same for the different cases. It may even be that none of the indicators used in classifying one case will be involved in classifying another.[42]

In conclusion it should be said that the difference between securing an expression of attitude through a life-history document and through the formal opinion questionnaire is, of course, purely a difference in the type of stimuli situation through which we elicit a verbal response. In the case of the life-history document it is assumed that these stimuli are more indirect, covert, subtle, and slantwise, and thus more disposed to secure an unbiased self-revelation, as compared with the more direct and pointed questionnaire of opinion. This is a doubtful contention for two reasons: (1) The questionnaire method of eliciting response is capable of the same subtleties, covertness and indirection as are the stimuli which elicit a life-history document. (2) It is certainly fallacious to assume

[42] *Ibid.*, pp. 818-19.

that even such documents as private letters and diaries are free from the pretenses, fantasies, and rationalizations that undoubtedly invalidate to a greater or lesser degree the face value of answers to a questionnaire. Says Kimball Young: "In truth, the digging up of past experiences by association is open to considerable qualification. In turn, personal documents, such as letters, memoirs, diaries, and confessional autobiographies are open to the same criticism that all forms of self-rating, self-revelation, and self-analysis are. The real motivations, especially, are not easy to get at from self-analysis, even though the main features of the mental mechanisms may be thus uncovered. If we wish to push the matter fully into the realm of motives, we may have to go beyond the ordinary autobiographical procedure. Here the interview, the method of mental analysis used by the psychiatrist, *and the objective check-up from other sources of information* come into play."[43] (Italics ours.) That the conditions which bias, warp, and invalidate the face value of the responses in the letter, diary, or life history probably will be *different* from those which achieve a similar effect in the case of the questionnaire, is undoubtedly true. But the assumed superiority of life-history documents, as such, over questionnaires, as such, is an erroneous notion which has probably arisen from the use of extremely crude forms of questionnaires with correspondingly crude results. The point is that the results of neither form of response can be taken at their face value. Both must be considered as responses to a total situation, and their face value as indications of subjective attitudes, or as indices of action in a real life situation, can be approached only by detailed analysis of the total situation under which the verbal responses are secured. No one questions the value of life-history documents, however, and they should be used for what they are worth. Perhaps their chief drawback is the difficulty of securing them in sufficient numbers and for a representative sample of a population. The greatest value of life-history documents probably lies in their capacity to suggest hypotheses for more intensive investigation.

[43] Kimball Young, "The Measurement of Personal and Social Traits," *Publ. Amer. Sociol. Soc.*, 21:103, 1927.

I. THE SOCIAL SURVEY

The social survey is not a special method of research distinct from the methods that have been discussed in preceding chapters. The survey is rather a type of organization of field work. That is, a social survey might employ any or all of the methods we have considered. There are no additional or distinctively "survey" methods. The term, "social survey" is conventionally used to designate (*a*) a more or less comprehensive inquiry into numerous aspects of a situation, (*b*) as it exists at a given time in a given community, (*c*) with rather definite educational, propagandist, and ameliorative purposes.[44] As such it may utilize any or all known methods of inquiry.

Historically considered, the practice of systematic inquiry regarding human resources and social conditions goes back to very early times. Thus, Herodotus records a survey of the population and wealth of Egypt about 3050 B.C. preparatory to the building of the pyramids. The Bible records censuses of population and taxable wealth. Such surveys were repeated by various Roman emperors and are illustrated in early England by William the Conqueror's preparation of the Domesday Book. Since the sixteenth century and the rise of national states a very large number of censuses and other inquiries on a large scale have been made at regular and irregular intervals.[45] The social survey in its more modern and restricted sense is usually thought of as beginning with the work of Le Play in his *Les Ouvriers Européens* (1855) or with Charles Booth's *Life and Labour of the People of London* (1891–1897). After that time, the latter type of inquiry became extremely popular, especially in the United States.[46] The ready availability of funds for such purposes

[44] *Cf.* E. W. Burgess, "Social Survey — A Field for Constructive Service by Departments of Sociology," *Amer. Jour. Sociol.*, 21:492, 1916. C. Aronovici, *The Social Survey*, p. 5. M. C. Elmer, *Social Research*, (Prentice-Hall, 1939), Chap. 11.

[45] For an outline of the development of such inquiries since 1500, see G. A. Lundberg, "Statistics in Modern Social Thought," in *Contemporary Social Theory*, edited by H. E. Barnes, H. and F. Becker, (Appleton-Century, 1940), Chap. 6.

[46] *A Bibliography of Social Surveys*, by A. Eaton and S. M. Harrison (Russell Sage Foundation, 1930), contains more than 2700 references to surveys completed up to January, 1928.

from the great research foundations doubtless contributed to this popularity and a vast number of good, bad and indifferent surveys resulted. The survey movement in its traditional form is perhaps already past the peak of its popularity and may be destined further to decline,[47] at least as far as surveys under private auspices are concerned. Among the reasons for this trend are (1) the rising standards of sociological inquiries, (2) the shrinkage of private research funds, (3) a more discriminating scrutiny of "survey" projects by the Foundations, and finally (4) the great extension of governmental records and inquiries. Some of the data usually secured with great trouble and expense through surveys will probably be provided (*a*) by a more comprehensive registration or regular collection of data by governmental agencies[48] and (*b*) by scientific inquiries into specific problems proceeding on definite hypotheses, and employing more refined methods of the types reviewed in previous chapters.

Since the form of inquiry known as the social survey is nothing but a coordinated series of investigations or researches into related problems, all the methods and principles of observing and recording data previously considered in this book may properly be applied to the survey. The practice of classifying the survey as a distinct method of research, or even placing it on a coordinate basis with such methods as the historical method, the case method, and the statistical method is clearly erroneous, as a survey may, and usually does, utilize all of these methods. It could, in fact, have no existence apart from these methods. The term "survey" merely indicates a certain degree of comprehensiveness and coordination with respect to the content or subject matter of an inquiry. A social survey, then, is a series of investigations or research projects undertaken with the purpose of emphasizing the relationship of the findings. It has no methods of its own, nor any techniques of gathering and

[47] *Cf.* Helen R. Jeter, "Research and Statistics in Social Work," *Social Work Year Book*, Russell Sage Foundation, 1941.

[48] Consider, for example, the inclusion of questions regarding housing and unemployment in the Federal Census of 1940; the large number of regional and local studies on all the conventional "social survey" topics by the Work Projects Administration; and such comprehensive surveys of the whole nation as are represented by *Recent Social Trends*, (McGraw-Hill, 1933).

evaluating data other than those outlined in Chapter IV and discussed in subsequent chapters. Only in the emphasis on coordination of individual investigations, and therefore in the organization of its field work, does it call for separate consideration.

1. General procedure in social surveys.—Since the only factor which distinguishes the social survey as a form of inquiry from other forms of investigation is its inclusiveness and the coordination of a number of individual research projects, the organization of the field work of the survey is the only aspect which concerns us in the present connection. Chapin has summarized the technique of organization for a survey as follows:

> The usual procedure is to get together a central committee representative of all the helpful and cooperative interests in the local community. An outside expert in survey work is then employed to plan and direct the investigation. This director usually finds it helpful to secure the cooperation of several state and national welfare organizations in carrying out the undertaking. A quick size-up of the local situation is then made by the directing expert. This preliminary stock-taking usually brings to the surface conditions of chronic as well as acute social maladjustment and points the way to further intensive study in special fields.[49]

The desirability of securing for field work the backing of a representative group of community agencies was discussed in an earlier section. Whether or not financial backing is sought, the approval of these agencies facilitates the field work and lays the foundation for community organization to carry out programs of social improvement suggested by the findings.

The need of skilled direction in social surveys cannot be too strongly emphasized. Too many reports show clearly the biased hand of the propagandist. Too frequently surveys are undertaken, not to reveal all the facts, but to accumulate facts to support the program of the individual or group sponsoring or conducting the inquiry. Sometimes this bias is not even intentional but merely the result of lack of training in scientific method on the part of the director. Everything which was said in Chapter II regarding bias, and the

[49] F. S. Chapin, *Field Work and Social Research,* (Century, 1920), p. 109.

need for intellectual honesty in observing first-hand data, applies with special force to the direction and conduct of social surveys.

After the scope of the survey has been determined in a general way, a quick size-up of the whole field by an expert is necessary as a basis for organizing the detailed field work. The general scope may vary from a city block in all its aspects to an entire region, with national or international boundaries. As the geographic scope increases, it is usually necessary to limit the scope of the content or subject-matter of the survey. Thus the most recent survey of Pittsburgh is devoted chiefly to "an examination of agencies and institutions that provide social and health services." [50] The scope of a survey both in content and geographic inclusiveness is determined in the last analysis by its purpose and the resources available for carrying it on. To determine how the available resources can best be utilized to achieve the ends sought is the purpose of the preliminary exploration by the director of the survey. On the basis of his preliminary exploration, he determines what sources to canvass and what situations to investigate. Only after this preliminary inquiry is it possible to organize the field work efficiently and economically.

2. Common defects of present social surveys. — The principal defect of social surveys up to the present time has been the lack of standardized procedure and the resulting incomparability of the data accumulated. Facts are of significance only in relation to other facts. Unless the data collected with reference to one community can be evaluated according to recognized standards, or compared to similar data in other communities, they are of doubtful scientific value and are unsafe as guides for social action. Too many surveys are nothing but collections of data on a variety of topics without any attempt to evaluate the facts with reference to recognized standards or to compare them with similar facts in other communities. If evaluation is attempted, it is too frequently on the basis of the surveyor's subjective judgment of what are reasonable standards, or to further the ends for which the "survey" was undertaken.

It is in connection with the need for comparison and

[50] P. Klein, *A Social Study of Pittsburgh,* (Columbia Univ. Press, 1938).

evaluation of the data collected that the standards formulated by state or national organizations in various fields become valuable guides. Some of these standards, as well as the standardized measuring-sticks developed in various fields for scoring institutions, have been discussed at length in Chapter IX. By means of such devices social surveying may become standardized, and the results have all the significance of accurate social measurement with reference to definite norms. With the development of a complete set of measuring instruments of the type considered in Chapter IX, social surveying would become mainly the technique of applying these measuring-sticks to the institutions and activities of a community and reporting the results in standard and comparable terms. This practice would vastly increase the value of social surveys. First, it would make it unnecessary for each surveyor to invent his own methods of measurement. We have considered in preceding chapters the difficulties involved in constructing measuring scales, and the need for thorough experimentation and revision before a reliable measure can be attained. The use by surveyors of their own untested devices is almost certain to result in conclusions of very dubious value. Second, the utilization of standard methods at once makes the findings of significance by enabling comparisons with recognized norms, and with other communities. By the development and use of standardized methods the social survey may become a scientific enterprise of significance to social research instead of a "fad and an industry." [51]

J. CONCLUSION

The development of social science will demand the gathering of data in the field to an increasing degree. The principles of sociology, economics, and political science cannot be deduced from armchair speculations nor from the ponderous philosophical tomes on human nature. The objective observation and recording of actual group behavior as it occurs is a constant necessity.

The collection of first-hand data on a large scale involves

[51] F. H. Giddings, *The Scientific Study of Human Society*, (Univ. of North Carolina Press, 1924), p. 185.

many problems in the selection, training, organization, and supervision of field workers. It involves also the development of a reliable and efficient technique of stimulating, observing, recording, and evaluating human responses, whether carried on in the more or less formal interview or otherwise. This technique is as yet somewhat undeveloped, but is becoming the object of such study and analysis as promises much for its improvement. Finally, the social survey has made its appearance in recognition of the need for factual social data as a basis for social adjustments. Most of the social surveys up to the present time, however, have suffered from the lack of standardized methods of social measurement, and a failure on the part of social surveyors to apply the better measuring devices which do exist. Being itself merely a coordinated series of investigations employing a variety of research techniques, the results of a social survey cannot, of course, have any greater reliability than the individual techniques employed are capable of securing. In short, the sporadic nature, the lack of standardized technique, and the lack of skilled direction of social surveys have made the data collected of doubtful scientific value.

Many of the questions which it is hoped can be answered by a social survey cannot in fact be answered by this type of inquiry. Some of the necessary data can be secured only through a more accurate, general, and standardized recording of certain social events as they occur. If there existed reliable and comparable records of all relevant social events and processes, such as now exist, for example, for births and deaths, social surveys for almost any social grouping could be conducted pretty much as an accountant checks the books of a corporation and reports on its status. In short, many social surveys are merely attempts to make up for inadequate community bookkeeping. Inadequate and unstandardized recording of social data is a major handicap both to practical administration and to the development of the social sciences. We turn, therefore, to this problem in the next chapter.

K. SUGGESTIONS FOR FURTHER STUDY

1. For a good brief summary of the history of modern social surveys, see Pauline Young, *Scientific Social Surveys and Research* (Prentice-Hall, 1939), Chaps. 1 and 2. See also her bibliography on this subject, pp. 535-41.

2. For a detailed consideration of interviewing, with special reference to social work, see Pauline Young, *Interviewing in Social Work,* (McGraw-Hill, 1935). A more general treatise is W. Van Dyke Bingham and B. V. Moore, *How to Interview,* (Harper, 1931), especially Chap. 6 ("The Interview in Marketing Studies and Commercial Surveys,") and Chap. 7 ("Interviews in Social Case Work"). Extensive bibliographies are to be found in this volume. Especially recommended as supplementary reading to the present chapter is *The Technique of Marketing Research,* (McGraw-Hill, 1937), Chaps. 3 and 4, by P. F. Lazarsfeld.

On the reliability and consistency of information secured through interviews as checked against other records, see Gertrude Bancroft, "Consistency of Information from Records and Interviews," *Jour. Amer. Statis. Asso.,* 35:377-81, 1940.

3. On the effects of bias on the part of the interviewer, see S. M. Harvey, "A Preliminary Investigation of the Interview," *Brit. Jour. Psychol.,* 28:263-87, 1938. Also S. A. Rice, "Contagious Bias in the Interview," *Amer. Jour. Sociol.,* 35:420-23, 1929.

4. For an excellent appraisal of the life-history technique, see H. Blumer, *Critiques of Research in the Social Sciences: I.* An appraisal of Thomas' and Znaniecki's *The Polish Peasant in Europe and America,* (Social Science Research Council, 1939). This publication includes also the report of a panel discussion of the subject by a number of social scientists from different fields. See especially the admirable summary by Read Bain, pp. 191-202. For an especially competent series of case studies and a careful consideration of the scientific problems involved, see Mirra Komarovsky, *The Unemployed Man and His Family,* (Institute of Social Research, New York, 1940). See especially the Appendix.

5. "An Outline for Writing a Case History," will be found in K. Young, *Personality, and Problems of Adjustment,*

(Crofts, 1940), Appendix. For a more detailed treatment, see John Dollard, *Criteria for the Life History,* (Yale University Press, 1935). For recent examples of the technique in current surveys, see John Dollard, *Caste and Class in a Southern Town,* (Yale University Press, 1938); Allison Davis and John Dollard, *Children of Bondage,* (American Council on Education, Washington, D. C., 1940).

6. On the "participant observer" technique see Florence R. Kluckhohn, "The Participant-Observer Technique in Small Communities," *Amer. Jour. Sociol.,* 46:331-43, 1940. For examples of the use of this method, see William F. Whyte, "Corner Boys: A Study of Clique Behavior," *Amer. Jour. Sociol.,* 46:647-64, 1941; Pauline Young, *Pilgrims of Russian-Town,* (University of Chicago Press, 1932); Nels Anderson, *The Hobo, The Sociology of the Homeless Man,* (University of Chicago Press, 1923); C. M. Arensberg and S. T. Kimball, *Family and Community in Ireland,* (Harvard University Press, 1940).

7. On case recording as practiced by social work agencies see Florence Hollis, *Social Case Work in Practice, Six Case Studies,* (Family Welfare Association of America, New York, 1939). Also Margaret C. Bristol, *Handbook on Social Case Recording,* (University of Chicago Press, 1936).

8. An exhaustive bibliography of social surveys completed up to January 1, 1928, will be found in A. Eaton and Shelby M. Harrison, *A Bibliography of Social Surveys,* (Russell Sage Foundation, 1930). The most adequate recent bibliography of various governmental surveys and research projects is the *Index to Research Projects,* Vols. I-III (Works Progress Administration, 1938–39) and *Catalogue of Research and Statistical Publications,* (Division of Social Research, Work Projects Administration, March, 1941). A valuable book suggesting the types of questions that surveys are frequently designed to answer is Joanna C. Colcord, *Your Community* (Russell Sage Foundation, Second Edition, 1941).

9. For excellent examples of recent surveys see: *The Social Survey of Merseyside,* edited by D. Caradog Jones, (University Press of Liverpool, Hodder and Stoughton, Ltd., London, 1934), 3 Vols; Calvin F. Schmid, *Social Saga of Two Cities, An Ecological and Statistical Study of Social Trends*

in Minneapolis and St. Paul, (Minneapolis Council of Social Agencies, 1937); Philip Klein, *A Social Study of Pittsburgh Community Problems and Social Services of Allegheny County,* (Columbia University Press, 1938).

10. For an experimental study of interviewer bias, see K. H. Baker and Frank Stanton, "Interviewer-Bias and the Recall of Incompletely Learned Material," *Sociometry,* 5:1942. (Forthcoming.)

CHAPTER XII

SOCIAL BOOKKEEPING

There are in the United States today 88 different agencies of the federal government collecting statistics about economic and social conditions. This amount appears to be a great deal to those who made out the 135 million returns on 4700 different forms in 1938, not a regular decennial census year. But a state cannot administer its affairs without knowledge of the conditions it deals with, any more than engineers can build the Grand Coulee dam without expert knowledge about steel and concrete. The engineer learns about his materials in the laboratory, while the government obtains its knowledge from statistics. In both cases the knowledge needs to be precise, the result of measurement. What if the census costs forty million dollars, it is necessary and worth it to the state, as well as to the uncounted users of census data. . . On the basis of the journal publications using statistics, the sciences are ordered as follows, psychology, economics, sociology, anthropology, political science, and history. That statistics is becoming a method of measurement in all these sciences is evidenced by the fact that during the past third of a century, the proportion of space used for formulae, tables and charts in these particular journals has increased 115 per cent. — *William F. Ogburn.*[1]

A. SOCIAL ORGANIZATION AND THE GATHERING OF SOCIAL DATA

If it is conceded that the advancement of the social sciences is dependent upon a more comprehensive collection of accurate and comparable social data, then the problem of providing for the systematic collection and filing of such data becomes of primary importance. The difficulties of controlled laboratory experimentation in the social sciences and the consequent inability to speed up artificially the particular social process which we wish to observe, make it all the more necessary to record social data as they occur over long periods of time. There are many vital types of social processes which cannot be reliably analyzed and interpreted except over periods of time embracing many generations. This characteristic of many social data, as well as the fact that we must have to some degree the cooperation of people in general for the collection of such data, makes us dependent to a large extent upon public authority for any comprehensive survey of social conditions and processes. The problem of the systematic gathering of social data on a large scale becomes, therefore, a problem of social organization. Hence,

[1] "Statistical Trends," *Jour. Amer. Statis. Asso.*, 35:256, 258 (Part II), 1940.

the problem of keeping the organization free from partisan and political influence is the same as in the case of organizations for other purposes. This becomes especially important in the case of an organization that is collecting data upon which social and political administration largely depend.

The necessity of keeping public records of social data has for various practical reasons been recognized since ancient times. Rulers and governments have always found some recording of vital, economic, and other social statistics indispensable to the guidance of public affairs. While these records have usually been regarded as useful chiefly from the standpoint of practical administration rather from the standpoint of data for science, these data have nevertheless become the basis for whatever generalizations underlie public policy. Not only is this true of the records kept by governments, but most institutions, public and private, have enormous archives of records. Taken together these constitute a collection of millions of facts which, if they were intelligible, would represent an important record of social experience from which scientific generalizations of great importance might be drawn. But unfortunately these records are largely useless on account of their incompleteness, inaccuracy, and incomparability. In short, because the principles of collecting data emphasized throughout this book have not been observed in the recording of the mass of data found in present-day government and institutional archives, these data are for the most part largely useless for scientific purposes.

B. THE DEVELOPMENT OF NATIONAL AND INTERNATIONAL STATISTICS

With the expansion of political units and the rise of national and international economic contacts and political organizations, the need for *uniformity* in methods of record-keeping has received increasing attention. Some of the older and stabler countries of Europe have partially comparable records extending over several centuries, not to mention the highly systematized methods of recording certain contemporary vital data. Not only has the need for uniformity and standardization within particular countries found ex-

pression, but since 1851 there have been comprehensive efforts to bring about international cooperation in securing standardized methods of recording vital social data. Willcox designates the London International Exhibition of 1851 as the beginning of international statistics. At this meeting, Quetelet suggested that an International Statistical Congress should soon be held for the purpose of securing greater comparability in statistical publications. This congress met in Brussels in 1853 under the presidency of Quetelet. It was followed by eight others during the following thirty-five years and by the establishment of a Permanent Commission to further international cooperative work in statistics. This series of meetings came to an end in 1878 and "somewhat the same international purpose was served thereafter by the International Congress of Hygiene and Demography which held its latest and perhaps its last session at Washington in 1912. At the Jubilee meeting of the Royal Statistical Society a new organization was started, modeled somewhat on the earlier Permanent Commission and with a limited professional membership. It met biennially, 1895–1913, and again in 1923."[2] In the meantime the International Statistical Institute at The Hague had carried on its work since 1885.[3] When the international meeting of 1923 took place, the need of the League of Nations for fuller and more trustworthy and comparable figures about the economic conditions of Europe lent especial significance to the systematic collection of certain data on an international scale. Recognizing the need for factual data as a basis for international as well as national government,[4] the League, in spite of its recent vicissitudes, still (June, 1941) publishes the *Monthly Bulletin of Statistics* and the *Statistical Year-Book of the League of Nations.* These publications contain comprehensive statistics for most nations on demography, production, labor, employment, cost of living, prices, finance and other subjects.

[2] Walter F. Wilcox, "The Relation of the United States to International Statistics," *Publ. Amer. Sociol. Soc.,* 18:18, 1923.

[3] Lucien March, "International Statistics and the League of Nations," *Publ. Amer. Statis. Asso.,* 17 (N. S.):635-38, 1920–21.

[4] For an account of the conditions leading up to the establishment of the League's statistical service, see "Report of the International Statistical Commission Appointed by the Council of the League of Nations," *Publ. Amer. Statis. Asso.,* 17 (N. S.):632, 1920–21.

C. THE UNITED STATES CENSUS BUREAU

The United States was represented at most of the international gatherings mentioned above. The first decennial census was taken in 1790, but the organization and scope of the work was rather rudimentary until 1850. A permanent Bureau of the Census was not established until 1902. While individual states, cities, and private organizations are, as we shall see, engaged in a great deal of more or less systematic collection of various kinds of data, the Federal Census Bureau remains our most important single agency for gathering, standardizing, and publishing statistical information. The present decennial census alone is an enormous undertaking which is constantly increasing its activities. We have already reviewed in a previous chapter (Chapter XI) the organization and conduct of a national census. The Census Bureau is also constantly engaged between the decennial censuses in the collection and analysis of vital statistics, agricultural, commercial, and other data. Among the more important of these activities may be mentioned: (*a*) special studies, such as the recent study on "Comparative Occupation Statistics, 1870 to 1930"; (*b*) preparation and revision of intercensal and post-censal population estimates; (*c*) a decennial census of religious bodies; and (*d*) the collection of data on mental patients, and statistics of criminal and judicial trends. This includes the establishment of a prison register to serve as a current inventory of the prison population, and a special sampling study of prisoners in the jails of the District of Columbia, Baltimore, and Philadelphia County. (*e*) In the field of vital statistics, the Bureau carries on extensive activities such as the following: (1) work toward a uniform vital statistics law for the various states and the use of standardized birth and death certificates; (2) revision of the International List of Causes of Death; (3) collection of hospital and institutional statistics; (4) collection of marriage and divorce statistics; (5) collection of accident statistics; and (6) special field projects for the improvement in registration of vital statistics. Extensive research is carried on in the Bureau for the improvement of census work and the development of

sampling techniques, upon which the further expansion of census work must perhaps be largely dependent.

Most of the data and results of these extensive activities are made available through numerous periodic publications, annual reports, and special reports and releases.[5] The most generally useful of these publications is perhaps the *Statistical Abstract of the United States,* published annually. This volume of about 1000 pages is perhaps the most valuable compilation of basically important social data to be found anywhere in the United States. The Census Bureau is likewise by far the most important organization in the United States, and perhaps in the world, for extending, improving, and standardizing the collection of the kind of social data upon which intelligent administration as well as comprehensive social research are increasingly dependent. The inquiry into housing conditions in the census of 1940, and the availability of census tract data from this census for some sixty cities will greatly increase the possibilities of refined and significant analysis of census statistics.

Extensive as are the activities of the Bureau of the Census, it is by no means (as we shall see below) the only agency, even of the federal government, engaged in the compilation of official statistics. For the year ending June 30, 1938, there were received some ten million income tax and informational returns and some seventeen million returns in connection with the old-age insurance program. Altogether about 97½ million administrative returns were made to the Federal government by individuals, institutions, and corporations. In addition about 38 million non-administrative returns were made during the same year, of which about 12 million were in connection with the census of unemployment. In a census year, the figure for non-administrative returns is, of course, greatly increased. The following account of the numbers of each type of return during a non-census year gives some idea of the subjects covered.[5a]

[5] For a partial list of these, see the *Report of the Secretary of Commerce,* 1940, pp. 57-58.

[5a] *Report of the Central Statistical Board,* (House Document No. 27, Government Printing Office, 1939), pp. 2, 7.

It is estimated that about 21,000,000 returns were filed for farms, making an average of approximately three returns per farm during the year. A total of 60,000,000 returns were received from industrial, commercial, and financial enterprises (not including farms but including all other types of business) or an average of about 20 per enterprise. On the average, each of these business returns contained answers to about 65 questions. These averages conceal wide differences among different classes of business. As might be expected, the returns, in general, were most extensive for regulated enterprises and for large-scale unregulated enterprises. The average number of returns filed per railroad was more than a thousand, and the average number of questions answered per return was about 150. For a group of large-scale enterprises, including manufacturers of automobiles, electrical apparatus, and steel, the average number of returns per respondent was about 250, while the average number of questions answered per return was 65. In contrast with these large concerns, the average independent retail store filed about 10 returns with the Federal Government and answered about 50 questions per return.[6]

If we add to this vast fund of information the reports collected by state and local governments, we gain some idea of the extent of government bookkeeping even in a country in which private enterprise is still supposed to be dominant. It is true, of course, that reports to the government have multiplied greatly during the past decade with the increased centralization of government and a rapid expansion of its functions. As Ogburn says: "Statisticians were called to Washington in hordes, and some who were not statisticians. The demand far exceeded the supply. From 1930 to 1938 the number of statisticians in government service in Washington is reported to have increased 700 per cent."[7]

D. TRENDS TOWARD INCREASING SCOPE AND STANDARDIZATION OF SOCIAL STATISTICS

In addition to the comprehensive activities of the Census Bureau and the administrative governmental agencies which have been reviewed above, recent years have also seen increasing attention to the collection and standardization of social statistics by other branches of the government and national associations of private agencies. We summarize be-

[6] *Ibid.*, p. 10.
[7] *Jour. Amer. Statis. Asso.*, 35:256, 1940.

low some of the more important of these developments during the decade 1930–40.

1. Uniform reporting of social work statistics. The need for comparable data regarding the work of the vast number of public and private agencies engaged in social service, relief, and other social work was recognized in this country at least as early as 1874.[8] A successful program toward that end, however, was not begun until 1924, by the Community Chests and Councils, Inc. The Russell Sage Foundation also promoted the cause by sponsoring a system of standardized voluntary reporting of such statistics, beginning in 1926. These undertakings led to the establishment in 1928 of a registration area for the collection of social work statistics. This work was at first under the auspices of a joint committee of the University of Chicago and the Association of Community Chests and Councils, but was soon transferred to the United States Children's Bureau. The service was further transferred to the Social Security Board in 1936. Since the establishment of this system of registration, forty-four cities have been brought into the registration area. The reporting covers the entire field of social work and also hospitals and out-patient departments. Financial as well as service data are included.

In addition to the data rendered available through the above system of registration, attention should be called to a number of handbooks published by various national agencies which define terms and describe statistical procedures. Such handbooks are now available for statistics of legal aid (1924), child guidance clinics (1930), public health nursing (1932), medical social service (1933), family welfare (1934), and hospitals (1935). A series of handbooks on similar and additional subjects was published by the New York State Department of Welfare during 1935 and 1936.

The *Social Work Year Book,* 1941, summarizes the present situation as follows:

> Despite the improvement both in quantity and in quality of the statistical reporting by social agencies since 1924, it is still impossible to produce for the United States a comprehensive statement

[8] See Emil Frankel, "Standardization of Social Statistics," *Soc. Forces,* 5:244-45, 1926.

of the volume and cost of public and private social services such as that published for the British Social Services by the Committee on Political and Economic Planning. Even in New York State, where the reporting is more complete than in any other state, there are gaps in reports from the voluntary social agencies in certain fields. Only in the field of public relief is it possible to publish each month for the United States as a whole, and for each state and local community, the number of cases, households, or persons receiving assistance and the amount of assistance received.

The present fairly complete state of public relief statistics is the result of cooperation among state and local agencies achieved under the leadership of five federal agencies (United States Children's Bureau, Federal Emergency Relief Administration—liquidated June 30, 1938, Work Projects Administration, Social Security Board, and Central Statistical Board) with the encouragement of the Joint Committee on Relief Statistics of the American Public Welfare Association and the American Statistical Association.[9]

2. The Children's Bureau. In addition to its early sponsorship of centralized reporting of social statistics as reviewed above, the Children's Bureau has been prominent in the development of systems of reporting many additional types of information. Extensive studies of maternal and infant deaths will be found among the publications of the Bureau. Its statistics of work permits issued to children of school age since 1920 provide an index of the decline of child labor. The collection and publication of standardized juvenile court statistics since 1927 is another contribution of the Bureau. The organization of nation-wide reporting of maternal and child health, crippled children, and child welfare services are among the Bureau's present projects.

3. The Federal Emergency Relief Administration and the Work Projects Administration. Statistics on public unemployment relief from 1933 to 1935 were collected with practical completeness by the Federal Emergency Relief Administration. Monthly reports from counties were secured through state offices covering case loads, relief granted, administrative costs, and sources of funds. With the withdrawal of federal grants for general relief in 1935, the reports from state and local units were unfortunately impaired. The

[9] *Social Work Year Book*, 1941, p. 486.

Works Progress Administration continued to publish monthly statistics, making its own estimates for areas for which reports were inadequate or lacking. This reporting system for general relief was transferred in 1937 to the Social Security Board. The W.P.A. continued to compile the statistics of the Works Program.

In addition to the above collection of statistics the Division of Social Research of this federal agency is constantly engaged in special studies, and has published a large number of excellent monographs, most of them based on extensive field studies. We have referred to some of these contributions in previous chapters.

4. The Social Security Board. The *Social Security Bulletin,* published monthly, is today the most important source on public relief statistics for the United States and for each state. The *Bulletin* not only continues the statistics of the W.P.A. mentioned above, but integrates them with other statistics of other public aid programs, and arrives at an estimated unduplicated count of households and persons receiving relief. The statistical series of private agencies transferred from the Children's Bureau is also continued by the Board. More detailed relief data are collected from public agencies in cities of 400,000 or more population. The *Social Security Bulletin* also publishes a variety of data and analyses bearing on its special functions.

5. The Bureau of Labor Statistics. This organization is the principal source of reliable data on a wide variety of subjects concerning employment, conditions of work, wages, and related subjects. These data are published in the *Monthly Labor Review,* which also publishes periodically a cost of living index number constructed from extensive reports of retail prices of different commodities throughout the nation.

6. The Central Statistical Board. — This Board was established by executive order in 1933, to coordinate, improve, and develop the vast statistical activities of the government. The Board has served in an advisory capacity in all the major developments in government statistics since that time. In 1938 the Board began a comprehensive study of the various returns made by the public to the government with a view of avoiding duplications and otherwise improving statistical

services. The Board recommended, among other more detailed suggestions, that although the collection of various types of statistics should be left for the most part to the administrative or other agencies chiefly concerned with the subject in question, this decentralization makes a strong statistical coordinating agency with adequate powers all the more desirable.[10] The functions of the Board were taken over, in 1940, by the Division of Statistical Standards of the Bureau of the Budget.

There are, of course, a large number of governmental as well as private organizations engaged in the collection of social statistics in their respective fields, which have not been mentioned above. A complete list would obviously include such agencies as the Office of Education, the United States Public Health Department, the Department of Agriculture, and the very large number of agencies dealing with strictly economic statistics. In the latter field, also, a very large number of private organizations such as the National Bureau of Economic Research and the National Industrial Conference Board should be mentioned. Even a brief review of all of these agencies, together with the large number of state and municipal organizations devoted to the recording of social experience in the form of statistics, is obviously beyond the scope of this chapter.[11] Our purpose has been not to provide a directory of statistical sources but to call attention to the principal movements under way for promoting uniform, standardized accumulation of such data.

E. OBSTACLES TO COMPREHENSIVE AND STANDARDIZED SOCIAL BOOKKEEPING

The adequate accumulation of social data for the guidance of society is, like most social problems, a problem of social organization. As such, the standardization of social statistics advocated in this chapter is retarded or prevented by many of the same obstacles which stand in the way of efficient social

[10] See *Report of the Central Statistical Board,* 1939, *op. cit.,* p. 30.

[11] An excellent example of what is being done in local urban areas along the lines here under discussion will be found in L. Wirth and Margaret Furez (Editors), *Local Community Fact Book,* 1938, (Chicago Recreation Commission, 1938). The studies by H. W. Green for the Cleveland Health Council are further examples of a very important type of social research.

organization in any field: tradition, inertia, and cultural lag. It is difficult to convince the great masses of people of the usefulness and necessity of the impersonal and abstract symbolic devices on which we must rely to an increasing degree for our control over our more removed social environments. The great majority of people are still organized on the basis of the simple face-to-face relationship of primary groups. As a result they observe and retain the significant data of past experience for the guidance of their future action by the informal and subjective method of casual observation and memory. The external storage of past experience in symbolic form — statistics and written records of every kind — which adjustment in the larger society requires, seems to the great masses of men to be an abstract, unreal, and unreliable process. The frequent inadequacy and unreliability of present records lends weight to the popular suspicion of statistics in all its forms. This in turn is largely due to the incompetence and carelessness of some persons at present engaged in recording and interpreting statistics. Comprehensive organization for the recording of social experience on a large scale such as the growing extent and complexity of our social contacts require is not an easy program to which to convert the general public.

In the first place, the increasing size of the task demands an increasing financial expenditure. With the public attitude what it is, more adequate provision for recording and bookkeeping in the public service and in private social agencies is likely to be the first to suffer from programs of economy. Even the Federal Census Bureau is severely handicapped financially in its efforts to tabulate and render useful many of the data which it collects. This is a situation which must to a large extent wait for its remedy on a change in public attitude. In the meantime the great private Foundations are performing a most valuable service in developing the technique of adequate recording as well as in educating social agencies and the public generally to the need and value of social bookkeeping.

In the second place, the divided jurisdiction of our system of government with its division of power between local, state, and national governments renders the adoption of uniform

methods of recording social data extremely difficult. The purposes for which a local community records certain data may not be as broad as the purposes for which state or national governments may desire these data. The data of the local community may therefore be inadequate for these larger purposes. Or, the data collected in different communities may be for the same purpose, but may be collected in different units and forms so as to render them incomparable. Only a central agency with mandatory powers such as the national Census Bureau can entirely remedy this situation. In the meantime, as we have seen, much excellent work has been accomplished in this direction through the voluntary cooperation of the states, local units, and private agencies with central agencies. Standardization on a national and international scale for many types of data is the goal to be sought.

F. SOCIAL CHANGE AND STANDARDIZED STATISTICS

Aside from the problems of social organization and authority which are involved in the standardization of social statistics, there are certain problems which arise from the changing social order itself. In our attempts to standardize units and methods of recording for long periods of time we are confronted with the fact that the changing social order renders certain units obsolete and demands the creation of new ones. Thus the census classification of occupations shows the disappearance of certain occupations and the appearance of others. Certain occupations have practically disappeared owing to mechanical inventions, and others have been absorbed by the rise of some new occupation. Or, the functions which used to be performed by people of certain occupational designation may have greatly altered, although the occupational name remains the same. For any or all of these reasons, a study of occupational classification of the present population and that of the middle of the last century might be very difficult and unreliable in its conclusions. The incomparability of statistics over a long period of time is a problem which will arise under the most carefully organized system of recording. New purposes which data are to serve, as well as new types of data, will dictate changes in units

from time to time. Under such circumstances the most we can do is either (1) to retain both the old and the new units and methods of recording if comparability with previous figures in the particular case under consideration is regarded as sufficiently important to warrant such duplication; or (2) to base the new classification, when possible, upon the old in such a way as to make the old and the new units subject to interpretation in terms of each other.

G. CONCLUSION

We have considered in this chapter some examples of the organizations which contemporary society has developed for the systematic recording of certain types of human behavior. These types of behavior have always been the object of informal observation and comment. Rough generalizations from these observations in the form of folklore have also come down to us from very ancient times. That is, age distributions, sex ratios, morbidity, mortality, marriage rates, and other characteristics of populations and uniformities of social behavior have always been a source of primary interest. It was perhaps also for this reason that these data were among the first to be formally and systematically recorded. In any case, it happens that on these subjects we have relatively reliable records, representing some millions of individual observations, extending in some countries over three centuries. We are so accustomed to the systematic collection of these demographic data that we take them more or less for granted.

As a result of the possession of these records we are able to demonstrate with great reliability the orderliness and predictability of certain demographic phenomena. This knowledge has in turn relieved us from the notions that these events are governed by personal, whimsical, or vindictive agencies of some sort. More especially has this type of knowledge enabled us to predict and make community provisions (*e.g.*, through insurance) for the events in question. The accurate prediction of the future size of populations, age distributions, marriage rates, etc., is obviously of basic importance in any social planning of such facilities as housing and educational and economic opportunities. Finally, it is in the

field of demography that we have made most progress toward the establishment of reliable generalizations of the type at which all science aims.

The superior development of positive, verifiable, sociological knowledge in the field of demography has erroneously led to the assumption that data of the type mentioned above are in some *intrinsic* sense more "objective" than it is possible for certain other data of social behavior to become. As we attempted to show in the first chapter, however, objectivity is not a quality or an attribute of data *per se*. Objectivity refers rather to the character of the symbols with which we communicate about phenomena. When these symbols have verifiably similar meanings to all men, we call the phenomena to which the symbols refer *objective*. It is not necessary to deny that such *objective designations* may be easier to invent for some phenomena than for others. To approach this question profitably, however, we would need to know the amount of time and effort devoted to such attempts for different types of phenomena throughout the centuries. It is sufficient for our purpose to point out that the advancement of science has consisted of gradually rendering "objective" increasingly large areas of traditionally "subjective" phenomena. We have merely assumed in this book that this process will continue also in the future and especially in the social sciences.

We have reviewed in preceding chapters some recent methods by which such "subjective" phenomena as attitudes, opinions, morale, and status have been objectively defined and measured. Likewise the extensive development of tests for the measurement and prediction of human intelligence, abilities, and achievement, represents steps toward the objectification of human characteristics that have been assumed to be "by nature" subjective and unmeasurable. There appears to be no reason why all of the above mentioned characteristics of populations and others should not be systematically recorded and take their place with other demographic data which are today regarded as necessary for intelligent social administration or scientific advance. It is only comparatively recently that statistics of illiteracy have made their

appearance among demographic data. Still more recently have we recognized that for many purposes "I Q" or "mental" age distributions are more important than the readily available chronological age. Such data, together with statistics on the distribution of wealth, migrations, rural-urban residence, etc., are already well recognized demographic data which we need to record as systematically as the traditional population, birth, death, and marriage statistics. The analysis of all of these data *over periods of time* and their correlation with each other and with other phenomena have yielded the most reliable and verifiable generalizations which the social sciences have thus far attained. When data of similar extent and reliability become available also for an unlimited variety of other characteristics and social behaviors of numerous different cultures, generalizations of demonstrable scientific validity will be more prevalent in the social sciences.[12]

The greatly increased registration of relatively standardized types of social data advocated above does not, of course, contemplate the abandonment or reduction of most of the special types of research considered through this book. On the contrary, a broader and more adequate foundation of basic data of the type we have here called demographic would greatly implement and stimulate the more academic and "purely scientific" research. The burden of collecting first hand the necessary basic data to test really important sociological hypotheses and theories is, as every researcher knows, frequently prohibitive. If the great extension of the statistical services which we have reviewed in this chapter seems itself to involve prohibitive amounts of time, personnel, and expense, three considerations should be remembered: (1) Increasing reliance upon scientific sampling will greatly reduce the size of such an undertaking as the decennial census without impairing its value. (2) The rapid development of statistical machinery easily makes possible the classification and more refined analysis of the amounts of data which would be

[12] A beginning toward a more comprehensive and systematic accumulation of data regarding different cultures is contemplated by the Cross-Cultural Survey under way at Yale University. See G. P. Murdock, "The Cross-Cultural Survey," *Amer. Sociol. Rev.*, 5:361-70, 1940.

quite unmanageable without such mechanical aids.[13] (3) As for the expense of a considerable extension and *improvement* of registration and gathering of standard social data, it is perhaps sufficient to point out that the census of 1940 in the United States cost much less than our latest battleship.

Finally, there will be those who are depressed by the "inhumanly dispassionate sweep of the scientific quest" and who will be impatient of the laborious and mundane methods proposed in this book. To those who feel this way we can only say with Veblen:

> One may approve or one may deprecate the fact that this opaque, materialistic interpretation of things pervades modern thinking. That is a question of taste, about which there is no disputing. The prevalence of this matter-of-fact inquiry is a feature of modern culture, and the attitude which critics take toward this phenomenon is chiefly significant as indicating how far their own habit of mind coincides with the enlightened common-sense of civilized mankind. It shows to what degree they are abreast of the advance of culture. Those in whom the savage predilection or the barbarian tradition is stronger than their habituation to civilized life will find that this dominant factor of modern life is perverse, if not calamitous; those whose habits of thought have been fully shaped by the machine process and scientific inquiry are likely to find it good. The modern western culture, with its core of matter-of-fact knowledge may be better or worse than some other cultural scheme, such as the classic Greek, the medieval Christian, the Hindu, or the Pueblo Indian. Seen in certain lights, tested by certain standards, it is doubtless better; by other standards, worse. But the fact remains that the current cultural scheme in its maturest growth, is of that complexion; its characteristic force lies in this matter-of-fact insight. Its highest discipline and its maturest aspirations are these.[14]

H. SUGGESTIONS FOR FURTHER STUDY

1. For a brief history of the development of social statis-

[13] "It is estimated that the following machines will be used for decennial census work: 907 manual key punch machines, 503 duplicating key punch machines, 12 alphabetic printing punch machines, 1 automatic multiplying punch machine, 11 alphabetic gang punch summary punch machines, 900 verifying machines, 92 sorting machines, 23 automatic comparing reproducers, 69 accounting machines, and 46 unit tabulating machines." *Report of the Secretary of Commerce*, 1940, p. 47.

[14] Thorstein Veblen, *The Place of Science in Modern Civilization*, (Viking Press, New York, 1932), pp. 29, 30.

tics, see George A. Lundberg, "Statistics in Modern Social Thought," (Chap. 6 in *Contemporary Social Theory,* edited by H. E. Barnes, H. and F. B. Becker, Appleton-Century, 1940). This chapter also contains references to the principal sources. See also "Proceedings of the Centenary Celebration, 1839–1939," *Jour. Amer. Statis. Asso.,* 35:213-308, 1940.

2. The best bibliographical reference and general digest of international population statistics and studies is the *Population Index,* published quarterly by the School of Public Affairs, Princeton University, and the Population Association of America, Inc.

3. For a report on attempts to coordinate official statistics in New York State, see R. L. Gillett, "Progress in State Statistics," *Jour. Amer. Statis. Asso.,* 35:335-40, 1940.

4. An excellent summary of the extensive statistical activities of the United States Government will be found in "The Statistical Services and Activities of the United States," in *The Inter-American Statistical Compendium,* (Central Statistical Board, Washington, D. C., May, 1940). See also *Statistical Activities of the American Nations,* edited by Elizabeth Phelps under the direction of the Inter-American Statistical Institute, Census Bureau Building, Washington, D. C., 1941. For research suggestions, see P. M. Hauser, "Research Possibilities in the United States Census," *Amer. Sociol. Rev.,* 6:463-70, 1941.

5. An excellent analysis of the type of cultural resistance which the scientific approach to human affairs is likely to encounter will be found in M. Goran, "The Literati Revolt Against Science," *Phil. of Science,* 7:379-84, 1940.

BIBLIOGRAPHY OF BIBLIOGRAPHIES

1. Arrington, Ruth E., *Time-Sampling Studies of Child Behavior,* Psychological Monographs Vol. 51, No. 2, 1939. (Amer. Psychol. Asso., Columbus, O.) Bibliography of 59 items on general and experimental observation of social behavior, chiefly in children.
2. Bain, Read, "Theory and Measurement of Attitudes and Opinions," *Psychol. Bull.,* 27:357-79, 1930. 261 titles.
3. Buros, O. K., *Educational, Psychological, and Personality Tests of* 1933–35, (Rutgers University, New Brunswick, N. J., 1936). See also G. Hildreth, *A Bibliography of Mental Tests,* (The Psychological Corporation, New York, 1939).
4. Crafts, L. W., Schneirla, T. C., Robinson, Elsa E., and Gilbert, R. W., *Recent Experiments in Psychology,* (McGraw-Hill, 1938).
5. Culver, Dorothy Campbell, *Methodology of Social Science Research: A Bibliography,* (Univ. of Calif., Berkeley, 1936). Annotated and classified titles.
6. Culver, Dorothy Campbell, *Bibliography of Crime and Criminal Justice,* (H. W. Wilson Co., N. Y., 1939). About 10,000 English bibliographical references annotated descriptively rather than critically.
7. Day, Daniel, "Methods in Attitude Research," *Amer. Sociol. Rev.,* 5:395-410, 1940. 200 readily available references in English covering the period 1937–39.
8. Eaton, Allen, and Harrison, Shelly, *Bibliography of Social Surveys,* (Russell Sage Foundation, 1928). Nearly 5,000 titles.
9. Eisenberg, P., and Lazarsfeld, P. F., "The Psychological Effects of Unemployment," *Psych. Bull.,* 35:358-90, 1938. 112 titles.
10. Franz, J. G., "Survey of Sociometric Techniques with an Annotated Bibliography," *Sociometry,* 2:76-90, 1939. 21 items.
11. Guilford, J. P., *Psychometric Methods,* (McGraw-Hill, 1936). Excellent bibliographies at the end of each chapter. See especially Chaps. 5-9 for material covered in the present text.
12. McCall, W. A., *Measurement,* (Macmillan, 1939), Chap. 7. Comprehensive list of tests and measuring instruments, mainly educational, pp. 91-133.
13. Psychological Corporation, The, *The Measurement of Artistic Abilities. A Survey of Scientific Studies in the*

Field of Graphic Arts, (The Psychological Corporation, N. Y., 1933). Available tests, studies of personality and temperament of artists and bibliography with summarized findings. 73 titles.

14. Quinn, J. A., "Topical Summary of Current Literature on Human Ecology," *Amer. Jour. Sociol.,* 46:191-226, 1940. 347 titles.
15. ——— *Interviewing and Case Recording,* (Russell Sage Foundation, N. Y., Bulletin No. 107, 1931). A selected bibliography.
16. Schettler, Clarence, "Topical Summaries of Current Literature : Personality Traits," *Amer. Jour. Sociol.,* 45:234-58, 1939. 308 titles.
17. Sewell, Wm. H., *The Construction and Standardization of a Scale for the Measurement of the Socio-economic Status of Oklahoma Farm Families,* (Oklahoma Agricultural and Mechanical College, Agricultural Experiment Station, Stillwater, Okla. Technical Bull. No. 9, 1940). The best annotated bibliography on measurement of socio-economic status. 77 titles.
18. Smith, B. O., *Measurement in Education,* (Bureau of Publications, Teachers College, Columbia University, New York, 1937).
19. U. S. Dept. of Interior, *Bibliography of Research Studies in Education* 1937–38, (Bull. No. 5, 1939, Office of Education, U. S. Dept. of Interior, Washington, D. C.).
20. Work Projects Administration, *Index to Research Projects,* Vols. I-III, 1938–39.
21. Work Projects Administration, *Catalogue of Research and Statistical Publications,* March, 1941.

AUTHOR INDEX

Allin, B. W., 309
Allport, F. H., 347, 348
Allport, G. W., 26, 129, 212, 213, 218, 252
Almack, R. B., 317, 319
Almack, R. N., 309
Alpert, H., xiv, 111, 347
Anderson, Nels, 395
Arensberg, C. M., 63, 347, 395
Aronovici, C., 388
Arrington, R. E., 64, 155, 328, 415
Ayres, C. E., 77

Bachelder, J. E., xiv
Bacon, F., 6
Bain, R., x, 16, 199, 212, 363, 394, 415
Baker, E., 235, 241
Baker, K. H., 396
Bancroft, G., 394
Barnes, H. E., 47, 116, 133, 320, 388, 413
Barnhart, E. N., 235
Bean, L., 246
Becker, F. B., 133, 388, 413
Becker, H., 116, 133, 388, 413
Beckman, E., 231
Bell, E. T., vi
Bentley, A. F., 79, 311
Bernard, J., 250
Bernard, L. L., ix, 12, 13, 46, 51, 53, 158, 212, 214, 219, 379, 380
Bierstedt, R., xiv
Bingham, W. V. D., 394
Bisbee, E., 13
Blair, M. M., 309
Blankenship, A. B., 195
Bloomfield, L., 81
Blumer, H., 246, 394
Boas, F., 115
Boettiger, L. A., 265
Bogardus, E. S., 233, 360, 363, 367
Bohr, N., 8
Booth, C., 388
Boring, E. G., 235
Bowley, A. L., 138, 140, 149, 150, 191
Bridgman, P. W., 13, 111
Bristol, M. C., 395
Brown, J. F., 3, 12, 347
Brown, L. O., 181
Buckley, H., 17
Buell, B., 347
Burdick, E. M., 294
Buros, O. K., 415
Burgess, E. W., 43, 384, 385, 388

Cajori, F., 8
Campbell, A. S., 231
Cantril, H., 2, 248
Cattell, J. M., 262, 263
Cavan, R. S., 200, 384
Chaddock, R. E., ix, 44, 120, 126, 146, 148, 151, 157
Chapin, C. V., 279
Chapin, F. S., ix, 7, 54, 56, 57, 67, 73, 79, 91, 92, 122, 127, 143, 144, 148, 154, 219, 252, 269, 288, 291, 292, 301, 307, 313, 315, 327, 330, 346, 347, 348, 390
Chapple, E. D., 63, 347, 368
Chase, S., 16, 17, 78, 81, 82
Chave, E. J., 238, 241, 252
Chesterton, G. K., 182
Christiansen, H. F., 67
Clapp, R. C., 86
Cohen, M. R., 250
Colcord, J. C., 309, 395
Comte, A., 116
Conklin, E. S., 239
Cook, S. W., 194
Cooley, C. H., 220, 247, 376
Copeland, M., 96
Corey, S. M., 247
Cottrell, L., 43, 384, 385
Cowden, D. J., 44, 157, 316
Crafts, L. W., 415
Criswell, J., 341, 346
Crow, C., 194
Croxton, F. E., 44, 157, 316
Culver, D. C., 415
Cunningham, E., 155
Curtis, A., 199

Darley, J. G., 231
Darwin, C., 52
Davidson, D. M., 339, 346
Davis, A., 395
Davis, Jerome, ix
Day, D., 231, 232, 246, 250, 251, 415
Dedrick, C. L., 158
Deming, W. E., 139, 157, 158
Dewey, J., vi, vii, 45, 79, 90, 133, 216
Dickson, W. J., 346
Dickson, W. M., 63
Diehl, N., 382
Dirac, 8
Dodd, S. C., xiv, 23, 30, 34, 43, 61, 101, 111, 154, 241, 270, 308, 312, 339, 344, 347
Dollard, J., 53, 381, 395

Drought, N., 248
DuBois, C., 278

Eaton, A., 388, 395, 415
Edgerton, H. A., 297
Eells, W. C., 324
Einstein, A., 2, 8, 102
Eisenberg, P., 415
Elmer, M. C., ix, 264, 268, 388
Esper, E. A., 81
Eubank, E. E., 85, 87

Faraday, M., 31
Faris, E., 220
Feldstein, M. J., 328, 346
Ferguson, L. W., 235, 239, 240, 246, 251
Firth, R., 372
Fisher, R. A., 157
Fiske, M., 372, 373
Ford, R. N., 231, 240, 241
Forsyth, C. H., 338
Fourier, 57
Frankel, Emil, 403
Franken, B., 226
Franz, J. F., 347, 415
Franzen, R., 189, 193
Freuchen, P., 375
Furez, M., 324, 406

Galileo, G., 160
Gallup, G., 158, 210, 221, 246
Garrett, H. E., 44
Gasnell, H., 246
Gaudet, H., 226, 229, 362
Ghiselli, E. E., 230
Giddings, F. H., vi, 104, 253, 392
Gilbert, R. W., 415
Gillette, J. M., v
Gillett, R. L., 413
Goldenweiser, A., 320
Goodenough, F., 155
Goodrich, C., 309
Goran, M., 413
Gould, K. M., 263
Green, H. W., 406
Guilford, J. P., 228, 235, 237, 252, 415
Gurnee, H., 235, 241

Haas, A., 8
Hansen, M. H., 139, 157, 158
Harrison, S. M., 388, 395, 415
Harrison, Tom, 371
Hart, A. B., 47
Hart, H., 111
Hartshorne, E. Y., 342
Harvey, S. M., 394
Hauser, P. M., 384
Hayes, M., 309
Heisenberg, W., 8
Hildreth, G., 252
Himes, N. E., xiv
Hobbes, T., 30, 51
Hogben, L., 82
Hollis, F., 395
Holmes, I., 158
Hooke, Robert, vi, 159
Horowitz, E. L., 66, 226
Horst, P., 44
Hotchkiss, G. B., 226
Hotelling, H., 384
Hull, C. L., 6, 9, 35
Huse, H. R., 207
Hutchinson, T. W., 15, 23

Inglis, Ruth, xiv

James, W., 80
Jenness, A., 348
Jennings, H. F., 335, 337, 339, 346, 347
Jeter, H. R., 390
Johnson, H. M., 250
Jones, D. C., 395

Katz, D., 214, 366
Keller, A. G., 47
Kelly, T., 384
Kendall, H. G., 138, 139, 147, 157
Kerstetter, L. M., 341
Kimball, S. T., 395
King, W. I., 98, 288
Kirkpatrick, C., 225, 248, 250, 252
Klein, P., 391, 396
Kluckhohn, C., 115
Kluckhohn, F. R., 376, 395
Komarovsky, M., 313, 394
Kozlowski, W. M., 13
Kreuger, E. T., 367

LaPiere, R. T., 213, 216, 218, 248
Lawsing, M., 378
Lazarsfeld, P. F., 66, 111, 133, 151, 158, 182, 209, 335, 347, 361, 365, 368, 372, 373, 384, 385, 394, 415
Leahy, A. M., 293, 297, 298, 299
Lenz, T. F., 225
Lenzen, V. F., 112
LePlay, F., 388
Lesser, A., 115
Lewin, K., 62, 312, 328, 347
Likert, R., 229, 232, 238, 239, 240, 241, 252
Lindsay, R. B., 111
Linné (Linnaeus), 109
Linton, R., 310, 327
Lippitt, R., 62
Lively, C. E., 309, 317, 319
Long, J. H., 294
Longmoor, E., 311, 324

Loomis, A. M., 64, 328
Loomis, C. P., 330, 339, 346
Lundberg, G. A., 43, 54, 79, 101, 106, 111, 132, 220, 249, 292, 311, 312, 313, 330, 332, 378, 388, 413
Lynd, H. M., 7, 43, 376
Lynd, R. S., 7, 43, 44, 376

McAtee, W. L., 44
McCall, W. A., 129, 415
McCandless, B., 204, 205, 207, 208
McCormick, M. J., 294
McCormick, T., 252
McKenzie, R. D., 315
McMillen, A. W., 86

Madge, C., 372
Malinowski, B., 372
Mandel, N. G., 67
Mangus, A. R., 317, 319
March, L., 399
Markey, J. F., x, 45, 220
Maxwell, C., 31
May, M., 302
Mayo, E., 63
Meltzer, H., 234
Mendeléeff, 109
Menefee, S., 65, 66, 194
Menger, K., 106, 111
Merton, R. K., 250
Mill, J. S., 1, 51
Miller, D. C., 7, 341, 346
Mills, 73
Mitchell, W., 23, 24, 27, 28, 29, 113
Monachesi, E., 43
Moore, B. V., 394
Moore, H. E., 315
Moreno, J. L., 62, 328, 329, 330, 335, 337, 346
Mosier, C. I., 295
Mowrer, E. R., 324
Murchison, C., 81, 212
Murdock, G. P., 9, 411
Murphy, G., 79, 229, 232, 238, 251, 252
Murphy, L., 79, 251

Nagel, E., 250
Nelson, E., 225
Newcomb, T., 251, 328, 346
Newstetter, W. I., 328, 346
Newton, I., 8, 11
Northway, M. L., 346

Odum, H. W., 315
Ogburn, W. F., 52, 76, 397, 402
Olson, W. C., 155
Owen, R., 57

Pace, C. R., 206, 231
Pallister, H., 227
Pareto, V., 12
Park, R. E., 383
Parten, M., 64, 158, 359
Patterson, D. G., 297
Pearl, R., 46
Pearson, K., 5, 11, 12, 15, 16, 28, 73, 246
Pearson, N., 246
Peck, L., 231
Pemberton, H. E., 229
Peters, C. C., 71, 252, 309, 320
Phelps, E., 413
Porterfield, A. L., 227

Quackenbush, O. F., 232
Quetelet, A., 399
Quinn, J. A., 415

Rae, S. F., 158, 210, 221, 246
Rankin, W. S., 279
Raskin, E., 194
Rasmussen, K., 375
Reicke, R., 8
Remmers, H. H., 248
Rice, S. A., 133, 223, 365, 394
Richmond, M. E., 361
Riley, J. W., 190
Ritchie, F. H. F., 169
Robinson, C. E., 246
Robinson, C. S., 195
Robinson, E. E., 415
Robinson, J. H., 116
Robinson, R., 347
Robinson, W. S., 385
Roehlisberger, F. J., 63, 346
Rollins, M. G., 208
Roper, E., 145, 151, 152, 158, 195, 206, 211, 349, 362
Rosander, A. C., 231, 252
Rubinow, I. M., 150
Rundquist, E. A., 91, 92, 252, 309, 341
Rutherford, E., 8, 17

Salisbury, P., 207
Sanderson, D., 111
Sandiford, P., 294
Sargent, J., 341
Sarle, C. F., 158
Scates, D. E., 251
Schettler, C., 416
Schilpp, P. A., vi, 45, 90
Schmid, C. F., x, xiv, 44, 324, 345, 395
Schmid, R. C., 252
Schneider, F., 279
Schneirla, T. C., 415
Schoenberg, E. H., 158, 359
Schrödinger, E., 8
Secrist, H., 126, 165
Sewell, Wm. H., 291, 294, 295, 296, 297, 298, 299, 307, 308, 416

Shaw, C. R., 311, 322
Shotwell, J., 116
Silance, E. B., 248
Sims, V. M., 294
Sletto, R. F., xiii, 91, 92, 184, 185, 189, 196, 197, 198, 203, 206, 209, 252, 295, 296, 309, 341
Small, A. W., 85
Smith, B. O., 416
Smith, J. H., 309
Smith, M., 200, 231
Solomon, R. S., 347
Sorokin, P., 220, 312
Spahr, W. E., 44, 133
Spencer, H., 12, 51, 79
Stagner, R., 194, 210, 224, 248
Stanton, F., 204, 205, 206, 207, 208, 396
Steele, M., 332
Stefansson, V., 193
Stephan, F. F., xiv, 136-137, 139, 140, 144, 150, 152, 153, 157, 158
Stevens, S. S., 111
Stouffer, S. A., 7, 158, 325, 345, 361, 384
Studenski, P., 195
Suchman, E. A., 204, 205, 207, 208
Sumner, F. C., 231
Sumner, W. G., vi
Sviatlovsky, E. E., 324
Swendsen, R. J., 44, 133
Sydenstricker, E., 61, 288
Symonds, P. M., 213

Thomas, Dorothy S., ix, 64, 328
Thomas, W. I., ix, 64, 246, 394
Thorndike, E. L., 196, 238, 252, 261, 298
Thurstone, L. L., ix, 212, 215, 236, 238, 239, 240, 241, 243, 244, 252, 295, 384
Tippett, 139
Toops, H. A., 202, 203
Tugwell, R. G., 74
Turner, R. C., 194

Veblen, T., 13, 14, 45, 412
Vold, G., 43

Wagner, I., 207
Wallis, W. D., 30, 94, 95
Watson, G. B., 224
Watson, R., 184, 185, 207
Webb, B., 40, 133
Webb, S., 40, 133
Westaway, F. W., 80, 109, 121
Wheeler, 109
Whipple, G. C., 176
White, M., 190
White, R. K., 62
Whitehead, T. N., 63
Whyte, W. F., 395
Wilcox, W. F., viii, ix, 399
Williamson, E. G., 231
Wilson, E. C., 362
Wilson, R. S., 382
Winslow, C. N., 231
Wirth, Louis, 324, 406
Wolf, A., 134

Young, E. F., 311, 324
Young, K., 129, 214, 387, 394, 380, 381
Young, P. V., 44, 324, 345, 360, 368, 394, 395
Yule, G. N., 138, 139, 147, 157

Zeleny, L. D., 63, 339, 341
Zeno, 81
Zipf, G. K., 309, 347
Znaniecki, F., 246, 394

SUBJECT INDEX

Abstracts, Social Science, 124-125
accuracy, degrees of, 42
adequacy, of samples, 149
ambiguity, test of, 244
ambiguous words in questions, 195-196
American Child Health Association, 280, 357
American Public Health Association, 255, 264, 279, 280
ammain, as a unit, 288
analogy, 118-119, 326
analysis and synthesis, 94, 95
answers to questionnaires, 205-208
anthropology, 115-116
appearance and reality, vi
appraisal form for health work, 279-288
appropriateness of units, 97
atom, social, 330
attitude measurement, Chap. 8; by general tests, 248-249; conditions influencing, 248-249; general considerations regarding, 246-251
attitudes, definition of, 212-214; importance of, 214, 219-220; measurability of, 215-217; measurement of, Chap. 8; methods of studying, 220-221; practical aspects of, 211-212; verbal, 217-220
attitudinal behavior, as index, 219-220
authority and backing of field worker, 350-352
autobiographies, 131

behavior, classified, 213; symbolic (see attitudes) verbal, 217-220
behavior, institutional (see institutional behavior)
bias and prejudice, nature of, 46-47; prevalence of, 47
bias, of interviewers, 365, 394
biographies, 220
bookkeeping, and public administration, 397-398; and social change, 408-409; community, Chap. 12, 393; national and international, 398-402; obstacles to, 406-408; social, Chap. 12
Brook Farm experiment, 57
Bureau of Labor Statistics (see *U. S. Bureau of Labor Statistics*)
business reply envelopes, use of, 185-186

candlepower, 84
cartographic methods of showing social relations, 343
case histories, 220, 379-388 (see also life history method)
case studies, 115; quantification of, 385-387
case study method, xiii
case work and quantitative methods, 25-27
cases, representatives of, 157
causation, 13, 14; and contingency, 73, 74
Census, Bureau of the, 400-402
census, Federal, field work of, 354-357; publicity for, 356-357; trial, 357
Central Statistical Board, 405-406
change (see social change)
Chapin scale, 289-290
cheating, 52
Child Welfare League of America, 264
Children's Bureau (see *U. S. Children's Bureau*)
city health work, scale for appraising, 279-288
clarity of units, 97-101
classification, Chap. 3, 10; examples of, 109; of methods of collecting data, 121-122; of occupations, 107-109; of questionnaires, 106; rules of, 106
coding, 174-175
Comanche social structure, 327
Committee on Conceptual Integration, 87
common-sense, and instruments, 300-302; fallacies regarding, 48-49
community bookkeeping, 393
comparability of units, 103-104
complexity, of social data, 16-18; meaning of, 18
concepts, tests of scientific, 87
configurations, social, 325-332
construction, of sociometric scales, 294-306
contingency, 73
correlation and explanation, 13, 14
corruption of scientific data, 52
critical ratio, method, of validation of items, 295
cross-cultural survey, 411

cultural regions, 317-320
curiosity, 31, 32

definitions, and instruments, 88; in the social sciences, 91-93; operational, 88-93
"degradation" of concepts, 92, 93, 304-325
demographic phenomena, 409-411
desires, 24, 25
diaries, 131, 221
diffused-symbolic institutions, 315
documentary sources (see sources)
"dummy" tables, 166-167

ecological maps, 317-323
ecology, Chap. 10, 310 ff.; techniques of, 315-325
enumerators, instructions to, 167-168
equal-appearing intervals, 235-238
errors, formulas for, 146; interviewer, 273; sampling, 272; schedule, 272; scorer, 273; seasonal, 272
ethics and science, 53, 54
evaluation, as measurement, 101
evaluative data, 101
experimental scales, 234-242
experimentation, and the statistical method, 65-74; as a level of scientific procedure, 7; difficulties of, 74-76; laboratory, 130; legislation as, 58-60; "natural," 56; possibility of social, 75; scientific social, 60-63; steps in, 7; techniques of, Chap. 2, 54-74
experiments, on relation of education to adjustment, 67-74; Utopian community, 57-78; on school children, 62, 63-65, 66
explanation, 13, 14

factors and weights, in sociometric scales, 260-265, 267-268, 277-278
Federal Emergency Relief Administration, 404
field investigation, methods of, 128-131
field staff, of Federal Census, 354-357; organization of, 352-359
field work, Chap. 11, 349 ff.; and community cooperation, 351-352; authority and backing for, 350-352; classification of, 349-350; for Federal census, 354-357; organization of, 352-359; refusals in, 358-359; 361-362; supervision of, 357-358
fields of science, 15-16
folk-designations, and scientific terms, 302-306
folklore and science, 77-78, 302-306
folk-terms and scientific concepts, 302-306, 345
follow-up letters, 203-205
Framingham experiment, 60, 61
friendship patterns, 314-315

gathering social data, Chap. 12; and social change, 408-409; and the advancement of science, 397; increase of, 397; obstacles to, 406-408; organization for, 397-402
generalization, 10; quantitative nature of, 132
geographers, and ecologists, 310-311
gradients in social phenomena, 322
Gresham's law, 12
group, as entity, 93-96; definitions of, 214; interviews, 371; structure, 326-333
guessing, 6, 216

handbooks of statistics, 403
health work, appraisal form of, 279-288
historical sources, 122
Hollerith tabulation, 174-175
horse-power, 84
hygienic behavior, measurement of, 270-278 (see also measurement of hygienic behavior)
hypothesis, 79; and theory, 119; dangers in use of, 120-121; definitions of, 120; of measurability of social phenomena, 217; working, 117-119

indices, 311-312, 317-320, 344-345; attitudinal behavior as, 218-220; importance of, 332, 334; in sociometric scales, 267; of adjustment, 341-342; of cohesion, 335-337; of interaction, 334; of metabolism or turnover, 342-343; of social status, 340-341; of sociation, 338-340; sociological, 332-343
institutional behavior, and attitudinal behavior, 306-307; demand for studies of, 253-254; measurement of, Chap. 9
institutions, nucleated and diffused — symbolic, 315
instruments, and common-sense, 300-302; development of, 49-51; for social measurement, 102-104; importance of, 33-34; schedules as, 159-160; types of, 50-51
"intangible" aspects of society, 307
intangibility of social phenomena, 19-21
intellectual honesty and the scientific method, 52-53

internal consistency, as measure of validity, 295-297
interpersonal relations, Chap. 10, 310 ff.; studies of, 328-333
interviewer, bias, 365, 394; errors, 200-201
interviewers, qualification of, 349, 353-354
interviewing, Chap. 11, 349 ff, 131, 359-371; appointments for, 362-363; favorable conditions for, 363; group, 371; panel, 371-375; participant observer, 375-379; principles of, 368-370; purposes of, 360; refusals, 358-359, 361-362; selecting persons for, 360-362; technique of, 364-371; telephone, 3
interviews, advantages of, 178-180; panel, 371-375; telephone, 213; types of, 367
introspection, 19-20, 331
intuition, 331
investigation, scope and purpose of, 162-165
irrelevance, criteria of, 243-245
isometric maps, 323
items, to include in schedule, 166; validation of, 294-297; weighting of, 297-300

judges of factors and weights in scales, 260-264, 267-268

knowledge, nature of, 45-46; purposes of, 31-34

laboratory experimentation (see experimentation)
language, as instrument of adjustment, 81; oral, limitations of, 83
law, scientific, 10; misconceptions regarding, 11-14; of falling bodies, 12-13
"leading" questions, 169, 193-194
League of Nations and statistics, 399
legislation, as experiment, 58-60
letters, 131; as sources, 221; soliciting replies to questionnaires, 183-186
library sources (see sources)
life history method, xiii, 131, 379-388; and questionnaire method, 386-387; relation of, to other methods, 380; standards of, 381
life stories, 220
Literary Digest poll, 145
literary methods, in social science, 29
literary samples, 157
literary studies of social phenomena, 42
machine tabulation, 174-175
machines, statistical, 411-412
maladjustment, 101
maps, as guides to adjustment, 81
Mass Observation, 371
mathematical symbolism, 313
mathematics and science, 83
measurability, 90
measurability, of attitudes, 215-217; of units, 102-103
measurement, as comparison, 90; as evaluation, 10; by general tests, 248-249; conditions influencing results of tests, 248; demand for, 253-254; obstacles to, 102-103; of attitudes, Chap. 8 (see also attitudes), 246-251; of hygienic behavior, 270-278; of institutional behavior, Chap. 9; of socio-economic status, 288-294; specifications of scale for, 270-275; techniques of construction of scale for, 275-278
measuring instruments, need for, 307-308
mental and physical, vi
methods, case study, xiii; complementary nature of different, 116-117; ecological, 315-325; historical and genetic, viii, 113-116; life history, xiii; need of composite, xiii, 113-114; non-quantitative, xiii; of collecting data, classification of, 121-122; of studying attitudes, 220-242 (see also attitudes); of weighting items in scales, 297-300; similarity of different, 113-114
mind, 21
mixed methods, of measuring attitudes, 238-242
morale, operational definitions, 91
motives, 24

National Board of Motion Picture Review, its use of panel technique, 372
National Child Labor Committee, 264
National Health Committee, 280
National Health Survey, 158
National Organization for Public Health Nursing, 280
National Recreation Association, 264
National Resources Committee, 125
National Tuberculosis Association, 280
natural areas, 320
"nature" of things, as sum total of associations, 100
New Harmony experiment, 57
non-ethical nature of science, 53-54
non-quantitative methods, xiii, 113

"normal," 100-101
novels, as samples, 157
nucleated institutions, 315
nucleation, social, 314; types of, 328-329

objectifying devices, development of, 49-51; types of, 50-51
objective and subjective, vi, 20
objective trends, 51
objectivity, nature of, 90; of data, 410-411
observation, 10; difficulties of, 45; random, 6
observer, participant, 375-379
occupational classification, problems of, 107-109
operational definitions, 88-93; and instruments, 88; in the social sciences, 91-93
opinions, definitions of, 212-214, 217 (see attitudes); measurement of, Chap. 8
oral language, limitations of, 83
order of merit scales, 235
orderliness, in social phenomena, 17
organization, charts, 326, 327; for gathering data, Chap. 12

paired comparisons, 234-235
panel, 371-375
participant observer, 375-379
pathology, 100
perfectionism, 41-42
personal investigation, advantages of, 178-180
physical and mental, vi
physical aspects of questionnaires, 170-171, 196-197
planning of research, 1, 31-39, 34-39, 162-165
polls, 145, 221-223
prediction, 20-22; from single case, 26-27, 115
prejudice and bias, nature of, 46-47; prevalence of, 47
pretesting questionnaires, 197-198; (see also testing)
probable error, 146-147; and representativeness of samples, 148
problems, selection of, 31-34
publicity for Federal census, 356-357
"pure" and "practical" research, 31-34
purpose of research, 162-165

qualitative and quantitative scales, 305-306; methods vs. quantitative, 23 (see also quantitative)
quantification of case studies, 385-387
quantitative methods, 22-25; and case work, 25-27, 115; generalization, 132; and literary methods, 29; and qualitative methods, 113; laboriousness of, 27-30 (see also qualitative)
questionnaires, Chap. 7; and life-history techniques compared, 386-387; and polls, 221-223; answers to, 205-208; arrangement of, 190; as special types of schedule, 182-183; circulation of, 202-203; classification of, 183; content and structure of, 186-190; cost of, 182; follow-up letters for, 203-205; integration of, 189; length of, 188-189; physical and esthetic aspects of, 196-197; pretesting of, 197-198; reliability of, 198-201; returns, interpretation of, 152-153; solicitation of replies to, 183-186; validity of, 201-202; waste questions in, 189-190; who answers, 205-208; wording of, 190-196
questions, categorized vs. "free," 191-193; corroboratory 201; definiteness of, 36; involving ambiguous words, 195-196; involving stereotyped words, 194-195; leading, 193-194, 169; subjective, 169; type of, 191-196

random sample (see sampling)
ranking scales, 225-227
rates, 315-317; in sociometric scales, 267
rating scales, 228-232
"real" and "apparent," vi; attitudes, 218
"reality," of certain symbolic representations, 311-313, of units, 95
recording of data, importance of, 76
refusals in interviewing, 358-359, 361-362
regional sampling, 154
regions, cultural, 317-319
registration of data, importance of, 393 (see bookkeeping)
relativity theory, 8
reliability, of questionnaires, 198-201; of samples, 145-149; of scale for measuring hygienic behavior, 272-273
religious factor, in community experiments, 58
representative samples (see sampling)
research budget, 38-39
research planning, 34-39 (see also planning)

research, planning of, 31-39; "pure" and "practical," 31-34; theory of, Chap. 1
research problems, relation of, 31-34
returns from questionnaires, 205-208
Russell Sage Foundation, and standardization of social statistics, 403-404

sampling, Chapter 5; by journalists and novelists, 157; cautions in, 149-153-157; conditions of valid, 135-137; in field work, 361-362; methods of, 137-145; mixed, 138-145; of time, 154-155; purposive, 138, 140-145; random, 136-140; representative, 148; regional, 154; size and reliability of, 145-149; steps in, 153-154; tests of adequacy of, 149
scales, arbitrary, 223-225; equal-appearing interval, 235-238; experimental, 234-242; for measurement of socio-economic status, 288-294; for city health work, 279-288; for measuring hygienic behavior (see measurement); mixed, 238-242; order of merit, 235; of social status, 289-294; paired comparison, 234-235; point, 224; qualitative and quantitative, 305-306; ranking, 225-227; rating, 228-232; sociometric (see sociometric scales); standardization of, 300-306; statement or unequal interval, 232-233; validation of, 300-306
schedules, Chapter 6; as instruments, 159-160; construction of, 165-175; filling of, 175-178; importance of, 160-162; physical aspects of, 170-171; questionnaire as, Chap. 7, 182-183; testing of, 175; types of, 162 ff; wording of, 167-170
science, vi, change in, vi
science, and folklore, 77-78; and theology, 77; definition of, 3-5; fields and methods of, 1-5, 15-16; levels of, 5-9; misconceptions regarding, 11-14; non-ethical nature of, 53-54; possibility of social, 14-16
science and mathematics, 83
scientific habit of thought, 1-2
scientific law, 10; misconceptions regarding, 11-14; nature of, 11-12
scientific method, 1; levels of, 5-9; steps in, 9-11
scientific procedure, levels of, 5-9
scientific terms and folk-designations, 302-306
scope and purpose of investigation, 162-165
"show me" test, 226-227
sigma method, 238-239; of weighting, 298
situation, 247; attitudes relative to, 218-220; response to total, 387
social atom, 330
social bookkeeping (see bookkeeping)
social change, and social statistics, 408-409
social data, gathering of, 397-398
social distance test, 233, 241, 242
social nuclei, 314-315
social organization and gathering of social data, 397-402
social patterns, 310 ff.
Social Science Abstracts, 124
social sciences, objective trends in, 51
Social Security Board, 405
social space, 312-313, 343
social status scale, 289-294
social survey (see survey)
Social Work Year Book 1941, 403-404
socio-economic status, defined, 288, 305-306; measurement of, 288-294
sociographic methods, 325-332
sociological indices, 332-343; importance of, 332-334
sociology, possibility of science of, 14-16
sociometric scales, availability of material for, 256-257; factors and weights in, 260-268, 277-278; general requisites of, 254-258; inclusiveness of, 257; judges of factors in, 260-264, 267-268; norms for, 256; problems in construction of, 258-259; reliability and validity of, 255-256, 268-269; scientific significance of, 259; steps in construction of, 294-306; summary of principles of construction of, 265-270; testing of, 269-270, 276, 278-279; types of factors in, 264-265; weighting of factors in, 257-258
solicitation of replies to questionnaires, 183-186
sources, critical evaluation of, 125-128; documentary, 122-125; field, 122; historical, 122; literary, 123-125
space, social, 312-313
standard error formula, 146-147
standardization of social statistics, 402-406; and social change, 408-409; obstacles to, 406-408; of scales, 300-306
statement scales, 232-233
statistical machines, importance of, 411-412

statistical method and experimentation, 65-74
statistics, agencies collecting, 403-406; and social change, 408-409; and the League of Nations, 399; handbooks of, 403; national and international, 398-399; obstacles to gathering, 406-408; trends toward increasing scope of, 402-406; uniform reporting of, 403-404
stereotyped words, influence of, 194-195
stereotypes, 224-225; mistaken for sample, 157
structure of groups, 326-333
"subconscious," the, 331
subjective and objective, vi, 20
subjective judgments, inadequacy of, 47-48
subjectivity, 90, 410-411; of social phenomena, 19-21
survey, Chap. 11, 349 ff; common defects of, 391-392; cross cultural, 411; general procedure in, 390-391; history of, 388; social, 388-392; trends in, 389
symbolic representation, importance of, 343-344; types of, 310-315

tabulation, machine, 174-175
techniques, importance of, vii; of interviewing (see interviewing)
telephone interviews, bias in, 203
terminology, Chap. 3; confusion of, 83-86; criteria of scientific, 86-87; importance of, 80-83
terms, tests of scientific, 87
testing of scales, 269-270, 276, 278, 279
theological thoughtways, 2
theology and science, 77
theory of research, 1, Chap. 1
time sampling, 154-155
tools, importance of, vii, schedules as, 159-160
"true" attitudes, 218
understanding and complexity, 18
unequal interval scales, 232-233
units, Chap. 3; appropriateness of, 97; as constructs of convenience, 93-96; clarity of, 97-101; comparability of, 103-104; measurability of, 102-103; "natural," 99-100; nature of, 95-96; requirements of, 96-104
United States Census Bureau (see Census Bureau)
U. S. Bureau of Labor Statistics, 405
U. S. Children's Bureau, 264, 404-405
U. S. Public Health Association, 280
U. S. Resettlement Administration, 330
Utopian Community experiments, 57-58

validation, critical ratio method of, 295; internal consistency method, 295-297; of items in scales, 294-297; of scales, 300-306
validity, 300-306; of hygienic behavior scale, 270-272; of questionnaire, 201-202; of tests, 242-245
values in sociology, 101
verbal behavior, 217-220

weighting of items in scales, 297-300
weights in sociometric scales, 260-265, 267-268, 277-278
What do scales measure?, 301-306
whole-part relationship, 93-96
"will," 21
wishes, 24-25
Woman's Home Companion, its use of panel technique, 372
wording of questionnaires, 190-196; of schedules, 167-170
working hypotheses, 9, 117-119; dangers in use of, 120-121; definition of, 120
Work Projects Administration, 125, 317, 395, 416; schedules of, 180-181

Date Due

RESERVE
RESERVE
(C) Ch. & Soc. Inst
RESERVE
RESERVE
Long - Church
RESERVE
RESERVE
(Long) Church in City & Industry
NO 9 '49
DE 21 '49
JA 4 '50
JA 19 '50
FE 2 '50
JUN 15 1986

HN29 .L96
Social research, a study in methods of
Princeton Theological Seminary-Speer Library
1 1012 00138 7176